

[illegible]

XX Novel purified corn tassel-derived polynucleotide useful for
 PT determining altered gene expression, to recover regulatory elements and
 PT to follow inheritance of desirable characteristics through hybrid
 PT breeding programs

PS Claim 1; SEQ ID 112; 201bp; English.

XX The present sequence describes a purified corn tassel-derived
 CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
 CC selected from those given in ABL/0627 to ABL/6833. The cdps sequences
 CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)
 CC can be used for determining altered gene expression, to recover
 CC regulatory elements and to follow inheritance of desirable
 CC characteristics through hybrid breeding programs. (I) are also useful
 CC in the evaluation, and alteration of desired characteristics associated
 CC with growth and development, disease resistance, environmental
 CC adaptability, quality and yield, and as molecular markers for studying
 CC inheritance of multigene traits in a plant breeding program. (I) can be
 CC used to produce a tassel-specific profile of gene transcription, a
 CC transcript image, to clone regulatory elements for use in transformation
 CC vectors, to express a polypeptide, to identify, isolate or extend
 CC identical or related corn tassel nucleic acid sequences from DNA
 CC libraries, in nucleic acid hybridization or amplification technologies,
 CC as query sequences to determine homology of known sequences, as probe
 CC for use in Southern or Northern hybridisation, and to identify the
 CC presence of and/or to determine the degree of similarity between two
 CC (or more) nucleic acid sequences.

SQ Sequence 340 BP; 102 A; 52 C; 84 G; 89 T; 13 other;

Query Match

Best Local Similarity 68.0%; Score 20.4; DB 24; Length 340;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 1 TGGACTCAGCTCGTCATCTCACCCTTCT 30
 126 TGGACACGCTACATGTCATCTCATTCTTCT -97

RESULT 2

ID AAC76076 standard; cDNA; 1323 BP.

AC AAC76076;

DT 08-FEB-2001 (first entry)

DE Human ORF1631 polynucleotide sequence SEQ ID NO:3261.

XX Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic; antihypertensive;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergic; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW thrombosis; contraceptive; ss.

XX Homo sapiens.

PN WO200056473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CUNAGEN CORP.

XX Shinkets RA, Leach M;

DR WPI, 2000-602362/57.

DR P-PSDB; AAB41867.

PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease

PS Claim 5; Page 2462; 5507bp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antidiabetic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antirheumatic; antithyroid; antianaemic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antineoplastic disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SQ Sequence 1323 BP; 239 A; 435 C; 343 G; 304 T; 2 other;

Query Match

Best Local Similarity 68.0%; Score 20.4; DB 21; Length 1323;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 1 TGGACTCAGCTCGTCATCTCACCCTTCT 30
 997 TGGCTCAGCTCGTCATCTCACCCTTCT 1026

RESULT 3

ID AAH90080 standard; cDNA; 1449 BP.

AC AAH90080;

DT 01-OCT-2001 (first entry)

DE Human bone marrow cDNA, SEQ ID NO: 324.

XX Human; bone marrow; antineoplastic; cytostatic; neuroprotective;
 KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
 KW immunosuppressive; gene therapy; cytokine cell proliferation;
 KW cell differentiation modulator; immune disorder; infection; cancer;
 KW human immunodeficiency virus; HIV; autoimmune disorder; hemophilia, ss.

XX Homo sapiens.

PN WO200153453-A2.

PD 26-JUL-2001.

PF 23-DEC-2000; 2000WO-US34960.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 30-NOV-2000; 2000US-0250583.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Drmanac RT;
XX
DR WPI: 2001-488707/53.
DR P-PSDB; AAM00961.
XX
PT Novel bone-marrow-expressed polynucleotides and polypeptides, useful
PT for treating e.g. cancer and immune deficiency disorders -
XX
PS Claim 1; Page 429; 648pp; English.
XX
CC The present sequence is one of 251 novel human polynucleotides
CC expressed in the bone marrow. The polynucleotide and the
CC polypeptide encoded by it are useful in the treatment of various
CC immune deficiencies and disorders. The deficiencies and disorders may
CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
CC infection, or may result from an autoimmune disorder, a coagulation
CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
CC suppression of an inflammatory response or treatment of a nervous
CC system disorder such as Alzheimer's disease. Detection of the presence
CC or increased expression of the polynucleotide or the protein it
CC encodes is useful for the diagnosis and/or prognosis of one
CC or more types of cancer. The polynucleotide and polypeptide can be
CC used as nutritional sources or supplements and in the screening of
CC chemical compounds as potential drugs.
XX
SQ Sequence 1449 BP; 334 A; 394 C; 446 G; 274 T; 1 other;
Query Match 68.0%; Score 20.4; DB 22; Length 1449;
Best Local Similarity 80.0%; Pred. No. 22;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 TGGACTCAGTCTTGGTCACTCACCCTCT 30
DB 327 TGGTCTCACTCCTTGGACTTGTCACTTGT 298
RESULT 4
AAK68321
ID AAK68321 standard; DNA; 15809 BP.
XX
AC AAK68321;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23133.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0200515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 14-JUL-2000; 2000US-0217487.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.

```

PS Disclosure: SEQ ID NO 23133; 3071pp + Sequence Listing: English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (1)
CC amino acid sequences given in AAM82170 to AAM91921. (1) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (1)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (1) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (1) by expressing inactive proteins or to
CC supplement the patients own production of (1). Additionally, (1)
CC polynucleotides may be used to produce the secreted (1). (1)
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (1) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 15809: BP: 3658 A; 4263 C; 4050 G; 3838 T; 0 other:

Query Match 65.3%; Score 19.6; DB 22; Length 15809;
Best Local Similarity 84.6%; Pred. No. 78;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 GACTCAGTCCTTGTCATCCTCACCCT 28
DB 13701 GACTCAGTCCTTTGTCATCTGGCCTT 13726
|||||
|||||

RESULT 5
AAS05390/C
ID AAS05390 standard; DNA: 81940 BP.
XX
XX AAS05390;
XX
XX 24-OCT-2001 (first entry)
XX
XX Human titlin (connectin) gene sequence.
XX
XX Human; titlin; connectin; pickwick mutation; cardiac specific exon; N2B;
XX titlin-related disease; zebrafish; heart failure; heart disease; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 133..80913
XX FT /tag= a
XX FT /product= "titlin"
XX
XX PN
XX MO200151666-A1;
XX
XX PD 19-JUL-2001.
XX
XX 12-JAN-2001; 2001WO-US01212.
XX
XX 12-JAN-2000; 2000US-015787.
XX
XX (GEHO ) GEN HOSPITAL CORP.
XX
XX Fishman MC;
XX
XX WPI: 2001-451869/48.
XX
XX P-PSDB; AADU5396.
XX
XX
XX Determining if a subject has or is at risk of developing a
XX titlin-related disease or condition, particularly heart failures,
XX comprises detecting the presence of a mutation in the titlin gene
XX
XX

```

Determining if a subject has or is at risk of developing a titin-related disease or condition, particularly heart failures, comprises detecting the presence of a mutation in the titin gene.

Disclosure, Page 35-57; 114pp; English.

The present sequence encoding for human titin (also known as connecting

CC is described in an invention relating to a novel method for determining
CC whether a subject has or is at risk of developing a titin-related
CC disease or condition. The method comprises analysing a nucleic acid of
CC a sample from the subject and detecting the presence of a mutation
CC (e.g. the pickwick mutation in the cardiac specific exon N2B) in the
CC titin gene, which indicates that the subject has or is at risk of
CC developing a titin-related disease. The zebrafish which has a phenotype
CC similar to mammalian heart failure is used as a model. The method is
CC useful for detecting an increased likelihood of heart disease, such as
CC heart failure, in a patient, so that appropriate intervention can be
CC instituted before any symptoms occur. The method may also be used to
CC facilitate determination of etiology of an existing heart condition,
CC such as heart failure, to identify compounds that can be used to treat
CC or prevent heart conditions, in prenatal genetic screening, e.g. to
CC identify parents who may be carriers of a recessive titin mutation.
CC Compounds identified using the methods may be used to treat patients
CC that have or are at risk of developing heart disease, e.g. heart
CC failure.
CC
SQ Sequence 81940 BP; 26373 A; 17100 C; 18899 G; 19568 T; 0 other;
XX
XX
Query Match 65.3%; Score 19.6; DB 22; Length 81940;
Best Local Similarity 84.6%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 TGGACTCAGTCTTGTCATCTCACC 26
DB 21155 TGTACTGAGTCTTGTTATATCACC 21130
|||
RESULT 6
ABK64829/c
ID ABK64829 standard; DNA; 81940 BP.
XX
AC ABK64829;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human benign prostatic hyperplasia gene #724.
XX
KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200212440-A2.
XX
PD 14-FEB-2002.
XX
PF 07-AUG-2001; 2001WO-US24708.
XX
PR 07-AUG-2000; 2000US-223323P.
XX
PR 05-JUN-2001; 2001US-0873319.
XX
PA (GENE-) GENE LOGIC INC.
XX
PA (NIBS) JAPAN TOBACCO INC.
XX
PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;
XX
DR WPI; 2002-257476/30.
XX
XX Identifying drugs for and diagnosing benign prostatic hyperplasia, by
XX detecting expression levels of one or more genes in prostate cells from
XX patient that are differentially regulated compared to normal prostate
XX cells
XX
PS Disclosure; Page 405-429; 444pp; English.
XX
XX The invention relates to a method of diagnosing (I) the onset or
XX progression of benign prostatic hyperplasia (BPH), or screening (II) for
XX or identifying an agent that modulates the onset or progression of BPH.
XX The method is based on changes in gene expression in BPH tissue isolated
XX from patients exhibiting different clinical states of prostate
XX hyperplasia as compared to normal prostate tissue. (I) comprises

CC detecting the expression levels of one or more genes in prostate cells
CC from the subject that are differentially regulated compared to normal
CC prostate cells. (II) comprises preparing a first gene expression profile
CC of BPH cells or BPH-like cell population, exposing the cells to the
CC agent, preparing a second gene expression profile of the agent exposed
CC cells, and comparing the first and second gene expression profiles.
CC (I) is useful for diagnosing the onset or progression of BPH. (II) is
CC useful for identifying an agent that modulates the onset or progression
CC of BPH. The methods are useful to present information identifying
CC the expression level in a tissue or cells, by comparing the expression
CC level of genes given in the specification in the tissue or cells to the
CC level of expression of gene in the database, and displaying the
CC expression levels of at least one gene in the tissue or cell sample
CC compared to the expression level in BPH. Agents using (II) are useful for
CC treating BPH or prostate cancer. ABK64106-ABK64860 represent human
CC benign prostatic hyperplasia gene sequences of the invention.
XX
SQ Sequence 81940 BP; 26373 A; 17097 C; 18901 G; 19569 T; 0 other;
XX
XX
Query Match 65.3%; Score 19.6; DB 24; Length 81940;
Best Local Similarity 84.6%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 TGGACTCAGTCTTGTCATCTCACC 26
DB 21155 TGTACTGAGTCTTGTTATATCACC 21130
|||
RESULT 7
AAA00522
ID AAA00522 standard; cDNA; 300 BP.
XX
XX AAA00522;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:513.
XX
KW Human; colon cancer; tumour; diagnosis; gene expression product;
XX probe; detection; cancerous state; metastasis; identification;
XX breast cancer; oestrogen receptor-positive breast cancer; therapy;
XX oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9958675-A2.
XX
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99WO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
XX
PR 15-MAY-1998; 98US-0085537.
XX
PR 15-MAY-1998; 98US-0085696.
XX
PR 21-OCT-1998; 98US-0105234.
XX
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
XX
PA (HYSE) HYSEQ INC.
XX
PI Williams IT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
XX
PI Reinhard C, Giese R, Randazzo F, Kennedy GC, Pot D, Kassam A;
XX
PI Lanson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
XX
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
DR WPI; 2000-126369/11.
XX
XX Polynucleotide library used to determine cancerous states of mammalian
XX cells -
XX
PS Claim 1; Page 302; 1097pp; English.
XX
XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA


```

XX AC AA227915;
XX XX 20-DEC-1999 (first entry)
XX DE Canine B7-2 protein coding sequence.
XX XX
XX B7: CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
XX KM allergic reaction; infectious disease; tumor development; canine;
XX KM graft rejection; inflammation; arthritis; atopic dermatitis; ss.
XX OS
XX OS Canis familiaris.
XX PN WO9947558-A2.
XX PD 23-SEP-1999.
XX PF 19-MAR-1999; 99WO-US06187.
XX PR 19-MAR-1998; 98US-0078765.
XX PR 17-APR-1998; 98US-0062597.
XX PA (HESK-) HESKA CORP.
XX PI Slim G, Yang S, Sellins KS;
XX PI WPI: 1999-571822/48.
XX DR P-PSDB; AAY41076.
XX PT New isolated B7 and CTLA4 nucleic acids, used to develop products for
XX PT treating, e.g. autoimmune and atopic diseases
XX PS Claim 1; Page 102-103; 148pp; English.
XX XX
XX CC The invention provides B7 and CTLA4 (T cell costimulatory proteins)
XX CC encoding nucleic acid molecules from dogs and cats. The proteins can be
XX CC expressed by standard recombinant methodology. The nucleic acid molecules
XX CC and the encoded proteins can be used for preventing or treating diseases,
XX CC e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
XX CC development, graft rejection, inflammation, arthritis and atopic diseases
XX CC such as atopic dermatitis. They can be used in mammals such humans, dogs,
XX CC cats, cattle, sheep or pets. The products can also be used for detection,
XX CC diagnosis and drug screening.
XX SQ Sequence 987 BP; 315 A; 215 C; 204 G; 253 T; 0 other;

Query Match 64.7%; Score 19.4; DB 20; Length 987;
Best Local Similarity 79.3%; Pred. No. 57;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGACTCAGTCTTGTCATCTCACCCTTCT 30
DB | | | | | | | | | | | | | | | | | |
901 GACTCTTCTCTGCTGCTCCTCCTCTCTCT 873

RESULT 11
AA227916
ID AA227916 standard; DNA; 987 BP.
XX
XX AA227916;
XX XX
XX 20-DEC-1999 (first entry)
XX DE Complementary strand of canine B7-2 coding sequence.
XX XX
XX B7: CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
XX KM allergic reaction; infectious disease; tumor development; canine;
XX KM graft rejection; inflammation; arthritis; atopic dermatitis; ss.
XX OS
XX OS Canis familiaris.
XX PN WO9947558-A2.
XX PT

```

```

PD 23-SEP-1999.
XX PF 19-MAR-1999; 99WO-US06187.
XX PR 19-MAR-1998; 98US-0078765.
XX PR 17-APR-1998; 98US-0062597.
XX PA (HESK-) HESKA CORP.
XX PI Slim G, Yang S, Sellins KS;
XX PI WPI: 1999-571822/48.
XX DR P-PSDB; AAY41076.
XX PT New isolated B7 and CTLA4 nucleic acids, used to develop products for
XX PT treating, e.g. autoimmune and atopic diseases
XX PS Claim 1; Page 103-104; 148pp; English.
XX XX
XX CC The invention provides B7 and CTLA4 (T cell costimulatory proteins)
XX CC encoding nucleic acid molecules from dogs and cats. The proteins can be
XX CC expressed by standard recombinant methodology. The nucleic acid molecules
XX CC and the encoded proteins can be used for preventing or treating diseases,
XX CC e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
XX CC development, graft rejection, inflammation, arthritis and atopic diseases
XX CC such as atopic dermatitis. They can be used in mammals such humans, dogs,
XX CC cats, cattle, sheep or pets. The products can also be used for detection,
XX CC diagnosis and drug screening.
XX SQ Sequence 987 BP; 253 A; 204 C; 215 G; 315 T; 0 other;

Query Match 64.7%; Score 19.4; DB 20; Length 987;
Best Local Similarity 79.3%; Pred. No. 57;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGACTCAGTCTTGTCATCTCACCCTTCT 30
DB | | | | | | | | | | | | | | | | | |
87 GACTCTTCTCTGCTGCTCCTCCTCTCTCT 115

RESULT 12
AA227921/C
ID AA227921 standard; DNA; 1795 BP.
XX
XX AA227921;
XX XX
XX 20-DEC-1999 (first entry)
XX DE Canine B7-2S protein encoding DNA.
XX XX
XX B7: CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
XX KM allergic reaction; infectious disease; tumor development; canine;
XX KM graft rejection; inflammation; arthritis; atopic dermatitis; ss.
XX OS
XX OS Canis familiaris.
XX PN WO9947558-A2.
XX PD 23-SEP-1999.
XX PF 19-MAR-1999; 99WO-US06187.
XX PR 19-MAR-1998; 98US-0078765.
XX PR 17-APR-1998; 98US-0062597.
XX PA (HESK-) HESKA CORP.
XX PI Slim G, Yang S, Sellins KS;
XX PI WPI: 1999-571822/48.
XX DR P-PSDB; AAY41076.
XX PT New isolated B7 and CTLA4 nucleic acids, used to develop products for
XX PT treating, e.g. autoimmune and atopic diseases

```

XX Claim 1; Page 109-111; 148pp; English.
 PS
 CC The invention provides B7 and CTLA4 (T cell costimulatory proteins)
 CC encoding nucleic acid molecules from dogs and cats. The proteins can be
 CC expressed by standard recombinant methodology. The nucleic acid molecules
 CC and the encoded proteins can be used for preventing or treating diseases,
 CC e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
 CC development, graft rejection, inflammation, arthritic and atopic diseases
 CC such as atopic dermatitis. They can be used in mammals such humans, dogs,
 CC cats, cattle, sheep or pets. The products can also be used for detection,
 CC diagnosis and drug screening.
 XX
 SQ Sequence 1795 BP; 592 A; 366 C; 347 G; 490 T; 0 other;

Query Match
 Best Local Similarity 64.7%; Score 19.4; DB 20; Length 1795;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 OY 2 GGACTCAGTCTTGGTGCATCTCCTCT 30
 DB 760 GTACCTTCTTCTGCTGCTCCTCTCT 732

RESULT 13
 AA227922
 ID AA227922 standard; DNA; 1795 BP.
 AC AA227922;
 DT 20-DEC-1999 (first entry)
 XX
 DE Canine B7-2S gene complementary DNA sequence.
 XX
 KW B7; CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
 KW allergic reaction; infectious disease; tumor development; canine;
 KW graft rejection; inflammation; arthritis; atopic dermatitis; ss.
 XX
 OS Canis familiaris.
 XX
 PN WO947558-A2.
 PD 23-SEP-1999.
 PF 19-MAR-1999; 99WO-US06187.
 PR 19-MAR-1998; 98US-0078765.
 PR 17-APR-1998; 98US-0062597.
 XX
 PA (HESK-) HESKA CORP.
 XX
 PI Slim G, Yang S, Sellins KS;
 DR WPI, 1999-571822/48.
 XX
 PT New isolated B7 and CTLA4 nucleic acids, used to develop products for
 PT treating; e.g. autoimmune and atopic diseases
 XX
 PS Claim 1; Page 112-114; 148pp; English.

XX The invention provides B7 and CTLA4 (T cell costimulatory proteins)
 CC encoding nucleic acid molecules from dogs and cats. The proteins can be
 CC expressed by standard recombinant methodology. The nucleic acid molecules
 CC and the encoded proteins can be used for preventing or treating diseases,
 CC e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
 CC development, graft rejection, inflammation, arthritic and atopic diseases
 CC such as atopic dermatitis. They can be used in mammals such humans, dogs,
 CC cats, cattle, sheep or pets. The products can also be used for detection,
 CC diagnosis and drug screening.
 XX
 SQ Sequence 1795 BP; 490 A; 347 C; 366 G; 592 T; 0 other;

Query Match
 64.7%; Score 19.4; DB 20; Length 1795;

Best Local Similarity 79.3%; Pred. No. 64;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 OY 2 GGACTCAGTCTTGGTGCATCTCCTCT 30
 DB 1036 GTACCTTCTTCTGCTGCTCCTCTCT 1064

RESULT 14
 AA227913/C
 ID AA227913 standard; DNA; 1897 BP.
 AC AA227913;
 DT 20-DEC-1999 (first entry)
 XX
 DE Canine B7-2 protein encoding DNA.
 XX
 KW B7; CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
 KW allergic reaction; infectious disease; tumor development; canine;
 KW graft rejection; inflammation; arthritis; atopic dermatitis; ss.
 XX
 OS Canis familiaris.
 XX
 PN WO947558-A2.
 PD 23-SEP-1999.
 PF 19-MAR-1999; 99WO-US06187.
 PR 19-MAR-1998; 98US-0078765.
 PR 17-APR-1998; 98US-0062597.
 XX
 PA (HESK-) HESKA CORP.
 XX
 PI Slim G, Yang S, Sellins KS;
 DR WPI, 1999-571822/48.
 DR P-PSDB; MAY41076.
 XX
 PT New isolated B7 and CTLA4 nucleic acids, used to develop products for
 PT treating; e.g. autoimmune and atopic diseases
 XX
 PS Claim 1; Page 97-99; 148pp; English.

XX The invention provides B7 and CTLA4 (T cell costimulatory proteins)
 CC encoding nucleic acid molecules from dogs and cats. The proteins can be
 CC expressed by standard recombinant methodology. The nucleic acid molecules
 CC and the encoded proteins can be used for preventing or treating diseases,
 CC e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
 CC development, graft rejection, inflammation, arthritic and atopic diseases
 CC such as atopic dermatitis. They can be used in mammals such humans, dogs,
 CC cats, cattle, sheep or pets. The products can also be used for detection,
 CC diagnosis and drug screening.
 XX
 SQ Sequence 1897 BP; 585 A; 400 C; 383 G; 529 T; 0 other;

Query Match
 64.7%; Score 19.4; DB 20; Length 1897;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 OY 2 GGACTCAGTCTTGGTGCATCTCCTCT 30
 DB 906 GTACCTTCTTCTGCTGCTCCTCTCT 878

RESULT 15
 AA227914
 ID AA227914 standard; DNA; 1897 BP.
 AC AA227914;
 DT 20-DEC-1999 (first entry)

```

XX Canine B7-2 gene complementary DNA sequence.
DE
XX B7; CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
KW allergic reaction; infectious disease; tumor development; canine;
KW graft rejection; inflammation; arthritis; atopic dermatitis; ss.
XX
OS Canis familiaris.
XX
PN WO9947558-A2.
XX
PD 23-SEP-1999.
XX
PF 19-MAR-1999; 99WO-US06187.
XX
PR 19-MAR-1998; 98US-0078765.
PR 17-APR-1998; 98US-0062597.
XX
PA (HESK-) HESKA CORP.
XX
PI Sim G, Yang S, Sellins KS.
XX
DR WPI; 1999-571822/48.
XX
PT New isolated B7 and CTLA4 nucleic acids, used to develop products for
PT treating, e.g. autoimmune and atopic diseases
XX
PS Claim 1; Page 101-102; 148pp; English.
XX
XX The invention provides B7 and CTLA4 (T cell costimulatory proteins)
CC encoding nucleic acid molecules from dogs and cats. The proteins can be
CC expressed by standard recombinant methodology. The nucleic acid molecules
CC and the encoded proteins can be used for preventing or treating diseases,
CC e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
CC development, graft rejection, inflammation, arthritis and atopic diseases
CC such as atopic dermatitis. They can be used in mammals such humans, dogs,
CC cats, cattle, sheep or pets. The products can also be used for detection,
CC diagnosis and drug screening.
XX
SQ Sequence 1897 BP; 529 A; 383 C; 400 G; 585 T; 0 other;

```

Query Match 64.7%; Score 19.4; DB 20; Length 1897;
 Best Local Similarity 79.3%; Pred. No. 65;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

QY 2 GGACTCAGTCCTTGTCATCTCACCCTCT 30
   1 1111 11111111 1111 1111
DB 992 GTCCTCTTCCCTTGCTGCTGCATCCTTCT 1020

```

Search completed: July 10, 2003, 19:52:36
 Job time : 66.3034 secs

Gencore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 13.4157 Seconds
(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631B-4

Perfect score: 30

Sequence: 1 tggactcagtccttggtcattcacccttct 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	63.3	33	1	US-08-299-682-13
2	19	63.3	33	1	US-08-347-657-1
3	18.4	61.3	550	3	US-08-705-771-1
4	18.4	61.3	786	5	PCT-US95-08295-1
5	18.4	61.3	1260	4	US-09-081-689-3
6	18.4	61.3	1297	4	US-08-858-207A-66
7	18.4	61.3	1329	4	US-09-305-984-13
8	18.4	61.3	1329	4	US-09-073-541A-13
9	18.4	61.3	1526	4	US-09-081-689-13
10	18.4	61.3	1526	4	US-08-751-189-1
11	18.4	61.3	7881	2	US-09-060-836-1
12	18.4	61.3	7881	2	US-09-184-445-1
13	18.4	61.3	8900	4	US-09-305-984-25
14	18.4	61.3	8900	4	US-09-073-541A-25
15	18.4	61.3	28882	4	US-08-961-527-140
16	17.8	59.3	3757	2	US-09-016-366A-13
17	17.8	59.3	3757	2	US-08-978-404B-19
18	17.6	58.7	2101	2	US-08-860-150-1
19	17.6	58.7	2101	2	US-09-338-132-1
20	17.6	58.7	5406	1	US-07-813-593-3
21	17.6	58.7	5406	1	US-07-977-451-5
22	17.6	58.7	5406	1	US-07-946-507-3
23	17.6	58.7	5406	1	US-08-252-517-5
24	17.6	58.7	5406	1	US-07-906-397A-5
25	17.6	58.7	5406	1	US-08-601-891-5
26	17.6	58.7	5406	2	US-09-021-324-5
27	17.6	58.7	5406	2	PCT-US92-02750-7

28	17.6	58.7	5406	5	PCT-US92-05401-5	Sequence 5, Appl1
29	17.6	58.7	5406	5	PCT-US92-09893-5	Sequence 5, Appl1
30	17.6	58.7	5470	2	US-08-443-861-1	Sequence 1, Appl1
31	17.6	58.7	5470	4	US-08-193-829B-1	Sequence 1, Appl1
32	17.6	58.7	43804	4	US-09-171-461-1	Sequence 1, Appl1
33	17.4	58.0	2533	4	US-09-149-922-5	Sequence 5, Appl1
34	17.4	58.0	2589	6	5212286-1	Sequence 5, Appl1
35	17.4	58.0	4149	2	US-08-737-715-1	Sequence 1, Appl1
36	17.4	58.0	35081	2	US-08-752-760A-1	Sequence 1, Appl1
37	17.2	57.3	330	1	US-08-081-539-84	Sequence 84, Appl1
38	17.2	57.3	330	1	US-08-466-647-84	Sequence 84, Appl1
39	17.2	57.3	2347	5	PCT-US96-03965-1	Sequence 1, Appl1
40	17.2	57.3	3776	1	US-08-162-809-7	Sequence 7, Appl1
41	17.2	57.3	6585	3	US-08-746-111-4	Sequence 4, Appl1
42	17	56.7	999	1	US-08-416-478A-4	Sequence 4, Appl1
43	17	56.7	999	2	US-08-474-988B-4	Sequence 4, Appl1
44	17	56.7	999	2	US-08-394-442B-4	Sequence 4, Appl1
45	16.8	56.0	426	4	US-09-712-016-32	Sequence 32, Appl1

ALIGNMENTS

RESULT 1
US-08-299-682-13/c
Sequence 13, Application US/08299682

Patent No. 5491063

GENERAL INFORMATION:

APPLICANT: Fisher, Mary Ellen

APPLICANT: Watson, Robert Malcom

TITLE OF INVENTION: Methods for In-Solution Quenching of

TITLE OF INVENTION: Fluorescently Labeled Oligonucleotide Probes

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/299,682

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Petry, Douglas A.

REGISTRATION NUMBER: 35,321

REFERENCE/DOCKET NUMBER: 9002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 814-2974

TELEFAX: (510) 814-2977

INFORMATION FOR SEQ. ID NO. 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 33 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-299-682-13

Query Match 63.3% Score 19; DB 1; Length 33;
Best Local Similarity 96.8% Pred. No. 9.2;
Matches 30; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
DB 31 tggactcagtccttggtcattcacccttct 1

RESULT 2

US-08-347-657-1/c

Sequence 1, Application US/08347657

Patent No. 5571673

GENERAL INFORMATION:

APPLICANT: Picone, Teresa

TITLE OF INVENTION: Methods for In-Solution Quenching of
Fluorescently Labeled Oligonucleotide Probes

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: U.S.A.

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/347,657

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Petty, Douglas A.

REGISTRATION NUMBER: 35,321

REFERENCE/DOCKET NUMBER: 9037

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 814-2974

TELEFAX: (510) 814-2974

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 33 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-347-657-1

Query Match

Best Local Similarity 63.3%; Score 19; DB 1; Length 33;

Matches 30; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db

1 TGACTCAGCTCTTGGTCATCTCAGCTTCT 30

31 TGACTCAGCTCTTGGTCATCTCAGCTTCT 1

RESULT 3

US-08-705-771-1/c

Sequence 1, Application US/08705771

Patent No. 6054289

GENERAL INFORMATION:

APPLICANT: Paul Moore, Reiner Gentz, Hongjin Ji,

TITLE OF INVENTION: Human Genes, Sequences and

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

ADDRESS: CECCHI, STEWART & OLSTEIN

STREET: 6 BECKER FARM ROAD

CITY: ROSELAND

STATE: NEW JERSEY

COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

US-08-705-771-1/c

Query Match

Best Local Similarity 61.3%; Score 18.4; DB 5; Length 786;

Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db

3 GACTCAGCTCTTGGTCATCTCAGCTTCT 30

205 GACCCCTCTTGGTCATCTCAGCTTCT 178

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/705,771

FILING DATE: August 30, 1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: MULLINS, J.G.

REGISTRATION NUMBER: 33,073

REFERENCE/DOCKET NUMBER: 325800-346 (PFI96)

TELECOMMUNICATION INFORMATION:

TELEPHONE: 973-994-1700

TELEFAX: 973-994-1744

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 550 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-705-771-1

Query Match

Best Local Similarity 61.3%; Score 18.4; DB 3; Length 550;

Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db

3 GACTCAGCTCTTGGTCATCTCAGCTTCT 30

122 GACCCCTCTTGGTCATCTCAGCTTCT 95

RESULT 4

PCT-US95-08295-1/c

Sequence 1, Application PC/TUS9508295

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: BREAST SPECIFIC GENES AND PROTEINS

NUMBER OF SEQUENCES: 30

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/08295

FILING DATE: 30-JUN-1995

CLASSIFICATION:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 786 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

PCT-US95-08295-1

Query Match

Best Local Similarity 61.3%; Score 18.4; DB 5; Length 786;

Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db

3 GACTCAGCTCTTGGTCATCTCAGCTTCT 30

205 GACCCCTCTTGGTCATCTCAGCTTCT 178

RESULT 5

US-09-081-689-3/c

Sequence 3, Application US/09081689

GENERAL INFORMATION:

APPLICANT: Wallis, Nicola G.

APPLICANT: Zalacain, Magdalena

APPLICANT: Throup, John

APPLICANT: Biswas, Sanjoy

US-09-081-689-3/c

TITLE OF INVENTION: Histidine Kinase
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/081,689
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/048,347
FILING DATE: 30-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: G410009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1260 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-081-689-3

Query Match 61.3%; Score 18.4; DB 4; Length 1260;
Best Local Similarity 95.0%; Pred. No. 27;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 GTCCTGGTCATCTCACCCT 28
|||||
DB 792 GTCCTGGTCATCTCACCCT 773

RESULT 6
US-08-858-207A-66/c
Sequence 66, Application US/08858207A
Patent No. 6348328
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: No. 6348328el Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smitlinkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 1297 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-858-207A-66

Query Match 61.3%; Score 18.4; DB 4; Length 1297;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 GTCCTGGTCATCTCACCCT 28
|||||
DB 792 GTCCTGGTCATCTCACCCT 773

RESULT 7
US-09-305-984-13/c
Sequence 13, Application US/09305984B
Patent No. 6331407
GENERAL INFORMATION:
APPLICANT: Toumanen, Elaine
TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
FILE REFERENCE: 1340-1-016N1
CURRENT APPLICATION NUMBER: US/09/305,984B
CURRENT FILING DATE: 1999-05-05
EARLIER APPLICATION NUMBER: 60/084,399
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: 09/305,984
EARLIER FILING DATE: 1999-05-05
NUMBER OF SEQ ID NOS: 76
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 1329
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-305-984-13

Query Match 61.3%; Score 18.4; DB 4; Length 1329;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 GTCCTGGTCATCTCACCCT 28
|||||
DB 264 GTCCTGGTCATCTCACCCT 245

RESULT 8
US-09-073-541A-13/c
Sequence 13, Application US/09073541A
Patent No. 6448224
GENERAL INFORMATION:
APPLICANT: No. 6448224ak, Rodger
APPLICANT: Toumanen, Elaine
TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
FILE REFERENCE: 1340-1-016
CURRENT APPLICATION NUMBER: US/09/073,541A
CURRENT FILING DATE: 1998-05-06
NUMBER OF SEQ ID NOS: 46

SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 13
LENGTH: 1329
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-073-541A-13

Query Match 61.3%; Score 18.4; DB 4; Length 1329;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 GTCTGTGTCATCTACCTT 28
||||| |||||||
DB 264 GTCTGTGTCATCTACCTT 245

RESULT 9
US-09-081-689-1/c
Sequence 1, Application US/09081689
Patent No. 6165992

GENERAL INFORMATION:
APPLICANT: Mallis, Nicola G.
APPLICANT: Zalacain, Magdalena
APPLICANT: Throup, John
APPLICANT: Biswas, Sanjoy
TITLE OF INVENTION: Histidine Kinase
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/081,689
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/048,347
FILING DATE: 30-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T.
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GM10009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1526 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-09-081-689-1
Query Match 61.3%; Score 18.4; DB 4; Length 1526;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 GTCTGTGTCATCTACCTT 28
||||| |||||||
DB 461 GTCTGTGTCATCTACCTT 442

RESULT 10
US-08-751-189-1

Sequence 1, Application US/08751189
Patent No. 5919656

GENERAL INFORMATION:
APPLICANT: Harrington, Lea A.
APPLICANT: Robinson, Murray O.
TITLE OF INVENTION: No. 5919656el Genes Encoding Telomerase Protein
TITLE OF INVENTION: 1
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen, Inc.
STREET: 1840 De Havilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/751,189
FILING DATE: 15-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy A.
REGISTRATION NUMBER: 34,688
REFERENCE/DOCKET NUMBER: A-433
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7881 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

US-08-751-189-1
Query Match 61.3%; Score 18.4; DB 2; Length 7881;
Best Local Similarity 78.6%; Pred. No. 35;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGACTGAGCTGGTGCATCTACCTT 29
||||| |||||||
DB 5358 GGACTGAGCTGGTGCATCTACCTT 5385

RESULT 11

US-09-060-836-1
Sequence 1, Application US/09060836
Patent No. 5981707

GENERAL INFORMATION:
APPLICANT: Harrington, Lea A.
APPLICANT: Robinson, Murray O.
TITLE OF INVENTION: No. 5981707el Genes Encoding Telomerase Protein
TITLE OF INVENTION: 1
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen, Inc.
STREET: 1840 De Havilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/060,836
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/751,189
FILING DATE: 15-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy A.
REGISTRATION NUMBER: 34,688
REFERENCE/DOCKET NUMBER: A-433
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7881 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-060-836-1

Query Match 61.3%; Score 18.4; DB 2; Length 7881;
Best Local Similarity 78.6%; Pred. No. 35;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGACTCAGTCTTGTCATCTCACCCTTC 29
DB 5358 GGACACAGTCCGTGGGCGAGCTGGCCTTC 5385

RESULT 12
US-09-184-445-1
Sequence 1, Application US/09184445
Patent No. 6174703

GENERAL INFORMATION:
APPLICANT: Harrington, Lea A.
APPLICANT: Robinson, Murray O.
TITLE OF INVENTION: No. 6174703el Genes Encoding Telomerase Protein
TITLE OF INVENTION: 1
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen, Inc.
STREET: 1840 De Havilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/184,445
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/751,189
FILING DATE: 15-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy A.
REGISTRATION NUMBER: 34,688
REFERENCE/DOCKET NUMBER: A-433
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7881 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-184-445-1

Query Match 61.3%; Score 18.4; DB 4; Length 7881;
Best Local Similarity 78.6%; Pred. No. 35;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGACTCAGTCTTGTCATCTCACCCTTC 29
DB 5358 GGACACAGTCCGTGGGCGAGCTGGCCTTC 5385

RESULT 13
US-09-305-984-25/c
Sequence 25, Application US/09305984B
Patent No. 6331407
GENERAL INFORMATION:
APPLICANT: Toumanen, Elaine
TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
FILE REFERENCE: 1340-1-016N1
CURRENT FILING DATE: 1999-05-05
EARLIER APPLICATION NUMBER: 60/084,399
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: 09/305,984
EARLIER FILING DATE: 1999-05-05
NUMBER OF SEQ ID NOS: 76
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 25
LENGTH: 8900
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-305-984-25

Query Match 61.3%; Score 18.4; DB 4; Length 8900;
Best Local Similarity 95.0%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 GTCCCTGGTCATCTCACCCTT 28
DB 7586 GTCCCTGGTCATCTCACCCTT 7567

RESULT 14
US-09-073-541A-25/c
Sequence 25, Application US/09073541A
Patent No. 6448224
GENERAL INFORMATION:
APPLICANT: No. 6448224ak, Rodger
APPLICANT: Toumanen, Elaine
TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
FILE REFERENCE: 1340-1-016
CURRENT APPLICATION NUMBER: US/09/073,541A
CURRENT FILING DATE: 1998-05-06
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 25
LENGTH: 8900
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-073-541A-25

Query Match 61.3%; Score 18.4; DB 4; Length 8900;
Best Local Similarity 95.0%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 GTCCCTGGTCATCTCACCCTT 28
DB 7586 GTCCCTGGTCATCTCACCCTT 7567

RESULT 15
US-08-961-527-140
Sequence 140, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue

Fri Jul 11 15:29:34 2003

us-10-087-631b-4.rni

Page 6

CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527

FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 140:
SEQUENCE CHARACTERISTICS:

LENGTH: 28882 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
US-08-961-527-140

Query Match 61.3%; Score 18.4; DB 4; Length 28882;
Best Local Similarity 95.0%; Pred. NO. 42;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 GTCTTGTCATCTCACCTT 28
|||||
DB 2314 GTCTTGTCATCTCACCTT 2333

Search completed: July 10, 2003, 20:27:14
Job time : 15.4157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 285.951 Seconds

(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631B-5

Perfect score: 33

Sequence: 1 gccacatgagtcgaagcgctgtgtataccg 33

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_cm:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_to:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	69.7	84126	9 HS125N5	AL008728 Human DNA
2	23	69.7	156998	23 AL136099	AL136099 Homo sapi
3	21	63.6	178874	2 AC096936	AC096936 Rattus no
4	20.8	63.0	1333	10 RMO251687	AI251687 Rattus no
5	20.8	63.0	187614	2 AC104398	AC104398 Rattus no
6	20.8	63.0	189192	2 AC119125	AC119125 Rattus no
7	20.8	63.0	205020	2 AC109903	AC109903 Rattus no
8	20.8	63.0	318696	2 AC102955	AC102955 Rattus no
9	20.4	61.8	3079	8 D89010	D89010 Aspergillus
10	20.2	61.2	24263	3 DMC118B3	AL031582 Drosophill
11	20.2	61.2	40396	2 AC087531	AC087531 Homo sapi
12	20.2	61.2	102227	2 AC020129	AC020129 Drosophill
13	20.2	61.2	106265	9 AC133294	AI133294 Human DNA
14	20.2	61.2	161668	3 AC104602	AC104602 Drosophill
15	20.2	61.2	179752	2 AC087522	AC087522 Homo sapi
16	20.2	61.2	188448	2 AC021368	AC021368 Homo sapi
17	20.2	61.2	203322	2 AC115149	AC115149 Rattus no
18	20.2	61.2	299970	3 AE003418	AE003418 Drosophill
19	20	60.6	696	14 E30279196	AI279196 Echovirus
20	20	60.6	744	14 E30279190	AI279190 Echovirus
21	20	60.6	744	14 E30279191	AI279191 Echovirus
22	20	60.6	744	14 E30279192	AI279192 Echovirus
23	20	60.6	3185	14 AF179613	AF179613 Echovirus
24	20	60.6	6489	14 HEC295172	AI295172 Human ech
25	20	60.6	80755	2 AC118767	AC118767 Rattus no
26	20	60.6	149967	2 AC025397	AC025397 Homo sapi
27	20	60.6	158138	2 AC129879	AC129879 Rattus no
28	20	60.6	161355	9 HSDJ60019	AL080314 Human DNA
29	19.8	60.0	771	6 AX057978	AX057978 Sequence
30	19.8	60.0	70113	2 AC101633	AC101633 Mus muscu
31	19.8	60.0	70113	2 AC101633	AC101633 Mus muscu
32	19.8	60.0	78132	2 AC022794	AC022794 Homo sapi
33	19.8	60.0	78132	2 AC022794	AC022794 Homo sapi
34	19.8	60.0	90150	9 AC068576	AC068576 Homo sapi
35	19.8	60.0	140757	9 AL157778	AL157778 Human DNA
36	19.8	60.0	154053	2 AC110331	AC110331 Rattus no
37	19.8	60.0	159507	9 AC004985	AC004985 Homo sapi
38	19.8	60.0	164165	2 AC103385	AC103385 Mus muscu
39	19.8	60.0	165413	9 AC078784	AC078784 Homo sapi
40	19.8	60.0	174327	2 AC068159	AC068159 Homo sapi
41	19.8	60.0	181561	9 AC015911	AC015911 Homo sapi
42	19.8	60.0	202308	2 AL772301	AL772301 Mus muscu
43	19.8	60.0	208752	2 AC124687	AC124687 Mus muscu
44	19.8	60.0	223088	2 AC099383	AC099383 Rattus no
45	19.8	60.0	239471	2 AC099284	AC099284 Rattus no

ALIGNMENTS

RESULT 1
HS125N5
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

HS125N5
Human DNA sequence from clone 125N5 on chromosome 6q26-27. Contains a putative novel gene, ESTs, STSS and GSSs, complete sequence.
AL008728.1 GI:3620992
HTG.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 84126)
Mashreqhi-Mohammadi, M.
Direct Submission

JOURNAL Submitted (04-NOV-1998) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk
 On Nov 2, 1998 this sequence version replaced gi:3717966.

COMMENT During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence is the entire insert of clone 125N5. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>
 125N5 is from the library RPL1 constructed at the Roswell Park Cancer Institute by the group of Plier de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcrpac2>

FEATURES

SOURCE

location/Qualifiers
 1. 84126
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="q26-27"
 /chromosome="6"
 /clone="RPL-125N5"
 /clone_id="RPL-1"
 68. 374
 /note="AluSx repeat: matches 1. 303 of consensus"
 1401. 1665
 /note="5 copies 53 mer 89% conserved"
 1642. 1770
 /note="3 copies 43 mer 81% conserved"
 join(2407. 2474,3240. 3472)
 /gene="dJ125N5.1"
 join(<2407. 2474,3240. >3472)
 /gene="dJ125N5.1"
 /product="dJ125N5.1 (PUTATIVE novel protein)"
 /note="match: EST AA732827"
 /evidence="not_experimental"
 2684. 2794
 /note="AluSg/x repeat: matches 178. 288 of consensus"
 complement(4097. 4389)
 /note="match: GSS A0053978"
 4980. 5072
 /note="MIR repeat: matches 75. 174 of consensus"
 5198. 5306
 /note="L2 repeat: matches 21. 136 of consensus"
 5628. 5762
 /note="L2 repeat: matches 554. 689 of consensus"
 7178. 7642
 /note="LTR repeat: matches 1. 450 of consensus"
 8264. 8481
 /note="HERVH repeat: matches 4314. 4531 of consensus"
 8575. 8676
 /note="HERVH repeat: matches 4170. 4272 of consensus"
 8679. 9122
 /note="HERVH repeat: matches 3186. 3637 of consensus"
 9105. 9287
 /note="HERVH repeat: matches 2669. 2847 of consensus"
 9293. 9405
 /note="MER4D repeat: matches 1. 114 of consensus"
 9408. 10330
 /note="L1PBI repeat: matches 5221. 6118 of consensus"
 10336. 10415
 /note="8 copies 10 mer aaaaagaag 69% conserved"
 10340. 10411
 /note="36 copies 2 mer aa 68% conserved"

repeat_region 10467. 10770
 /note="AluSg repeat: matches 1. 301 of consensus"
 repeat_region 10771. 11083
 /note="MER4B repeat: matches 206. 484 of consensus"
 repeat_region 11178. 11309
 /note="MER4D repeat: matches 865. 989 of consensus"
 repeat_region 11809. 12511
 /note="MER4D repeat: matches 236. 973 of consensus"
 misc_feature complement(12331. 12682)
 /note="match: GSS B30601"
 12691. 13063
 /note="match: GSS B51137"
 complement(12868. 13508)
 /note="match: GSS A0020723"
 14553. 14613
 /note="LTR16C repeat: matches 308. 373 of consensus"
 repeat_region 14650. 14763
 /note="L2 repeat: matches 2601. 2709 of consensus"
 14764. 15109
 /note="match: GSS A0032080"
 repeat_region 14881. 14922
 /note="MER5A repeat: matches 112. 157 of consensus"
 15110. 15197
 /note="2 copies 44 mer 86% conserved"
 15233. 15576
 /note="LTR2FB repeat: matches 1. 366 of consensus"
 16556. 16713
 /note="2 copies 79 mer 82% conserved"
 16733. 16843
 /note="3 copies 37 mer 83% conserved"
 19332. 19642
 /note="AluSg repeat: matches 1. 309 of consensus"
 19661. 19752
 /note="L1M4 repeat: matches 2964. 3055 of consensus"
 19753. 20067
 /note="AluSg repeat: matches 1. 312 of consensus"
 20068. 20298
 /note="L1M4 repeat: matches 2734. 2964 of consensus"
 20368. 20485
 /note="L1M4 repeat: matches 6249. 7977 of consensus"
 20571. 20721
 /note="MER57-internal repeat: matches 7145. 7305 of consensus"
 20848. 20895
 /note="MER57-internal repeat: matches 7387. 7434 of consensus"
 20932. 21206
 /note="L1PAB repeat: matches 5889. 6163 of consensus"
 21493. 21583
 /note="L1M4 repeat: matches 7492. 7586 of consensus"
 21860. 21962
 /note="L1M4 repeat: matches 7104. 7205 of consensus"
 22229. 22402
 /note="LTR1C repeat: matches 285. 461 of consensus"
 23093. 23196
 /note="HAL1 repeat: matches 680. 781 of consensus"
 23650. 23875
 /note="L1M6C repeat: matches 2385. 2295 of consensus"
 23949. 24038
 /note="LTR1C repeat: matches 1. 84 of consensus"
 26324. 26634
 /note="Alu repeat: matches 1. 311 of consensus"
 26811. 26895
 /note="L1M2 repeat: matches 6038. 6122 of consensus"
 complement(26912. 27500)
 /note="match: genomic DNA Z77936"
 27022. 27116
 /note="L1M1 repeat: matches 6064. 6159 of consensus"
 27124. 27180
 /note="3 copies 19 mer 91% conserved"
 27531. 27817
 /note="AluSg repeat: matches 6. 296 of consensus"
 28174. 28315
 repeat_region

```

repeat_region /note="MIR repeat: matches 44. 184 of consensus"
30062. 30361 /note="AlusX repeat: matches 1. 312 of consensus"
repeat_region 30864. 30935 /note="4 copies 18 mer 81% conserved"
repeat_region 31965. 32052 /note="2 copies 44 mer 86% conserved"
repeat_region 32326. 32382 /note="Charle3 repeat: matches 2644. 2700 of consensus"
repeat_region 33415. 33452 /note="MER63 repeat: matches 3. 38 of consensus"
misc_feature 33890. 34267 /note="match: GSS AQ178363"
repeat_region 34268. 34339 /note="2 copies 36 mer 94% conserved"
repeat_region 35047. 35550 /note="14 copies 36 mer 94% conserved"
repeat_region 35922. 36057 /note="MIR repeat: matches 39. 184 of consensus"
repeat_region 36297. 36364 /note="2 copies 34 mer 96% conserved"
repeat_region 37108. 37258 /note="L1P44 repeat: matches 6000. 6146 of consensus"
repeat_region 38408. 38501 /note="MIR repeat: matches 162. 258 of consensus"
prim_transcript complement(40975. 41312) /note="match: EST AA83825"
repeat_region 41592. 41711 /note="L2 repeat: matches 2290. 2418 of consensus"
prim_transcript complement(43225. 43547) /note="match: EST AA97687"
repeat_region 44041. 44093 /note="MIR repeat: matches 59. 113 of consensus"
repeat_region 44143. 44183 /note="MIR repeat: matches 218. 258 of consensus"
repeat_region 44214. 45044 /note="277 copies 3 mer tga 63% conserved"
repeat_region 44214. 45041 /note="46 copies 18 mer 63% conserved"
repeat_region 44215. 45042 /note="23 copies 36 mer 63% conserved"

Query Match 69.7% Score 23; DB 9; Length 84126;
Best Local Similarity 83.9% Pred. No. 10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCCACATGATGCGCAGCGCTGCTGATAC 31
DB 11528 GCCACCTGATGCGCAGCGCTGCGGAAC 11558

RESULT 2
LOCUS AL136099 156998 bp DNA linear HTG 16-AUG-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-270C4, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL136099
VERSION AL136099.10 GI:9796029
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 156998)
AUTHORS Phillimore,B.
TITLE Direct Submissison
JOURNAL Submitted (13-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Aug 11, 2000 this sequence version replaced gi:9588114.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC

```

```

Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba270C4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 15572 bases at least Q40
Consensus quality: 155690 bases at least Q20
Insert size: 156998; sum-of-contigs
Insert size: 157873; 2.1% error; agarose-fp
Quality coverage: 4.85x in Q20 bases; agarose-fp
coverage: 4.85x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
location/Qualifiers
source 1. 156998
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-270C4"
/clone_11b="RPC1-11.1"
1. 156998
/note="assembly_fragment:00254
clone_end:77
vector_side:right"

BASE COUNT 42309 a 35053 c 36018 g 43618 t
ORIGIN

Query Match 69.7% Score 23; DB 2; Length 156998;
Best Local Similarity 83.9% Pred. No. 10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCCACATGATGCGCAGCGCTGCTGATAC 31
DB 42942 GCCACCTGATGCGCAGCGCTGCGGAAC 42972

RESULT 3
LOCUS AC096936/C 178874 bp DNA linear HTG 13-JUL-2002
DEFINITION Rattus norvegicus clone CH230-41H3, *** SEQUENCING IN PROGRESS ***,
63 unordered pieces.
ACCESSION AC096936
VERSION AC096936.4 GI:21735373
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 178874)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsdrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Bardaria,J., Benton,J., Blinage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowle,S., Bileva,M., Brown,M., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,K.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
DeLaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinb,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frintz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Hernandez, O., Hodgson, A., Hughes, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J., Kovar, C.,
Kritovic, J., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lonsleged, H.,
Lodado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, A., Ma, J.,
Manshwar, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,
Miner, G., Mitchell, T., Mohabadi, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunolu, G.,
Ordunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojudo, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherrer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, T.,
Soderberg, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Swatek, A., Taylor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telitod, B., Thomas, N., Thomas, S.,
Umani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, O.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

Unpublished
2 (bases 1 to 178874)

Direct Submission
Submitted (03-OCT-2001)

Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 178874)

Worley, K.C.

Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:17973294.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GH0D

Center clone name: CH230-41H3

Sequencing vector: plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 113609 bases at least Q40

Consensus quality: 119045 bases at least Q30

Consensus quality: 124009 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
NOTE: This is a "working draft" sequence. It currently
consists of 63 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1
1169 contig of 1168 bp in length
1269 gap of unknown length
1269 contig of 1567 bp in length
2836 gap of unknown length
2935 gap of unknown length
4059 contig of 1123 bp in length
4158 gap of unknown length
5149 contig of 1356 bp in length
5515 gap of unknown length

5615
6703
6803
8109
8209
8210
9708
9709
9809
10889
10989
12003
12103
13623
13724
15455
15555
15556
17381
17481
19206
19306
20802
20902
22078
22078
22178
23587
23688
25662
25762
26938
27038
28300
28400
29951
30051
31787
31886
32898
32899
32999
34877
34877
36793
36893
39118
39219
40269
40369
42143
42243
43553
43554
44781
44881
47131
47231
49648
49748
50976
51076
53856
53955
55999
57099
60137
60237
62082
62182
64982
64989
65059
65060
67195
67295
69539
69540
72537
72538
72637
75265

6702: contig of 1088 bp in length
6802: gap of unknown length
8109: contig of 1307 bp in length
8209: gap of unknown length
9708: contig of 1499 bp in length
9808: gap of unknown length
10888: contig of 1080 bp in length
10988: gap of unknown length
12002: contig of 1014 bp in length
12102: gap of unknown length
13623: contig of 1521 bp in length
13724: gap of unknown length
15455: contig of 1732 bp in length
15555: gap of unknown length
17380: contig of 1825 bp in length
17480: gap of unknown length
19205: contig of 1725 bp in length
19305: gap of unknown length
20801: contig of 1496 bp in length
20901: gap of unknown length
22077: contig of 1176 bp in length
22177: gap of unknown length
23587: contig of 1410 bp in length
23687: gap of unknown length
25661: contig of 1974 bp in length
25761: gap of unknown length
26937: contig of 1176 bp in length
27037: gap of unknown length
28299: contig of 1262 bp in length
28399: gap of unknown length
29950: gap of unknown length
30050: contig of 1551 bp in length
31786: gap of unknown length
31886: contig of 1736 bp in length
32898: contig of 1012 bp in length
32998: gap of unknown length
34776: gap of unknown length
34876: contig of 1778 bp in length
36792: gap of unknown length
36892: contig of 1916 bp in length
39118: gap of unknown length
39218: contig of 2226 bp in length
40268: gap of unknown length
40368: contig of 1050 bp in length
42142: gap of unknown length
42242: contig of 1774 bp in length
43554: gap of unknown length
43554: contig of 1312 bp in length
44780: gap of unknown length
44780: contig of 1126 bp in length
47130: gap of unknown length
47230: contig of 2250 bp in length
49647: gap of unknown length
49747: contig of 2417 bp in length
50975: gap of unknown length
51075: contig of 1228 bp in length
53855: gap of unknown length
53955: contig of 2780 bp in length
55998: gap of unknown length
57098: contig of 3043 bp in length
60136: gap of unknown length
60236: contig of 3038 bp in length
62081: gap of unknown length
62181: contig of 1845 bp in length
64981: gap of unknown length
64989: contig of 2778 bp in length
65059: gap of unknown length
65060: contig of 2136 bp in length
67195: gap of unknown length
67295: gap of unknown length
69539: contig of 2244 bp in length
69540: gap of unknown length
72537: contig of 2898 bp in length
72637: gap of unknown length
75265: contig of 2628 bp in length


```

* 75266 75365: gap of unknown length
* 75366 78583: contig of 3218 bp in length
* 75367 78584 78683: gap of unknown length
* 78684 81964: contig of 3281 bp in length
* 81965 82064: gap of unknown length
* 82065 85985: contig of 3921 bp in length
* 85986 86085: gap of unknown length
* 86086 89524: contig of 3439 bp in length
* 89525 89624: gap of unknown length
* 89625 92377: contig of 2753 bp in length
* 92378 92477: gap of unknown length
* 92478 96946: contig of 4469 bp in length
* 96947 97047 100448: contig of 3402 bp in length
* 100449 100548: gap of unknown length
* 100549 102573: contig of 2025 bp in length
* 102574 102673: gap of unknown length
* 102674 106025: contig of 3352 bp in length
* 106026 106125: gap of unknown length
* 106126 109123: contig of 2998 bp in length
* 109124 109223: gap of unknown length
* 109224 113781: contig of 3558 bp in length
* 113782 112881: gap of unknown length
* 112882 116631: contig of 3750 bp in length
* 116632 116731: gap of unknown length
* 116732 120720: contig of 3989 bp in length

Query Match 63.6% Score 21; DB 2; Length 178874;
Best Local Similarity 82.8% Pred. No. 91;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GCCACATGATGTCGACGAGCGCTGTGTAT 29
Db 140041 GTCACATGACGTGCAAGAGCTGTGTCT 140013

RESULT 4
RNO251687 1333 bp mRNA linear ROD 24-FEB-2000
LOCUS Rattus norvegicus mRNA for pepsinogen F protein, strain SD.
ACCESSION AJ251687
VERSION AJ251687.1 GI:7105997
KEYWORDS pepsinogen F protein.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1333)
Moriyama, T., Ichinose, M., Tsukada-Kato, S., Omata, M., Narita, Y.,
Moriyama, A. and Yonezawa, S.
Molecular cloning of neonate/infant-specific pepsinogens from rat
stomach mucosa and their expressional change during development
Biochem. Biophys. Res. Commun. 267 (3), 806-812 (2000)
10673373
2 (bases 1 to 1333)
Kageyama, T.
Direct Submission
Submitted (25-NOV-1999) Kageyama T., Center for Human Evolutionary
Modeling Research, Primate Research Institute, Kyoto University,
Inuyama, 484-8506, JAPAN
location/Qualifiers
1. 1333
/organism="Rattus norvegicus"
/strain="SD"
/db_xref="taxon:10116"
/clone="prr723"
/dev_stage="4-day-old infant"
21. 1184
/function="unknown"
/codon_start=1
/product="pepsinogen F protein"
FEATURES
SOURCE
CDS

```

```

/protein_id="CAB75982.1"
/db_xref="GI:7105998"
/translation="MKMLWLVGIALVSECIKPIPLMKTKSMRENIRESHMKDYLEKY
PRRAHVLEDRNPSTVYEPNRNYLDLVYIGTISGTPPEFVYVLDGSSDLWPS
ITCSSPACAHKRVNVPVLOSSTFVSGRVNVAVSGESGLADIVKIGLVVAQA
FGLSLBEGREFMEHAFVDGILGIVPNIGLVGVVDNMLQGLIPQNLFAFLSK
DEKSVLMLGCVDPVSIVYHGLHMLWVPSVAGVSIWNGGIIICDGGCQIMDT
GISLVTPRSRSLIIONLIGAKASGDGEFLKCTINTLPITVITISVYVPVASAV
IRKHSNCRNRFESSTDDPEDPLWLGDFLALYTFVFRANNRIGLASAA"
sig_peptide 21..65
mat_peptide 66..1181
BASE COUNT 306 a 361 c 343 g 323 t
ORIGIN
Query Match 63.0% Score 20.8; DB 10; Length 1333;
Best Local Similarity 78.1% Pred. No. 1.1e+02;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 CCACATGATGTCGACGAGCGCTGTGTATACCG 33
Db 899 CCTATTGTGTCGACAGGCTTGTGTATGCGC 930

RESULT 5
AC104398 187614 bp DNA linear HTG 12-JUL-2002
LOCUS Rattus norvegicus clone CH230-128124, *** SEQUENCING IN PROGRESS
DEFINITION
AC104398.4 GI:21729960
VERSION AC104398.4
KEYWORDS HTG: HTGS. PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 187614)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Albrooks, S.L., Amaralunge, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbarta, J., Benton, J., Blmage, K., Blankenburg, K., Bonin, D.,
Bouck, J., Bowle, S., Brileva, M., Brown, M., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Datborne, S.R., David, R.,
Dayila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
DeLaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotco, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gottrell, J., Guenara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hughes, M., Holloway, C., Hollins, B.,
Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlssohn, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Deal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Ilean, C., Liu, J., Liu, W., Louised, H.,
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokkwo, S., Ogun, M., Okunolu, G.,
Oreguine, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I.,
Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,

```

TITLE
JOURNAL
AUTHORS
REFERENCE
TITLE
JOURNAL

COMMENT

Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. P., Zhong, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 187614)

Worley, K.C.
Direct Submission
Submitted (09-DEC-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 187614)

Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:18846151.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: G10D

Center clone name: CH230-128124

Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 115282 bases at least Q40

Consensus quality: 120980 bases at least Q30

Consensus quality: 125749 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).

NOTE: This is a 'working draft' sequence. It currently
consists of 72 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1025: contig of 1024 bp in length
1125: gap of unknown length
1125: contig of 1000 bp in length
2125: gap of unknown length
2225: contig of 1309 bp in length
3534: gap of unknown length
3634: contig of 1213 bp in length
4847: gap of unknown length
4946: contig of 1224 bp in length
6171: gap of unknown length
6270: contig of 1038 bp in length
7309: gap of unknown length
7409: contig of 1206 bp in length
8615: gap of unknown length
8715: contig of 1352 bp in length
10067: gap of unknown length
10167: contig of 1340 bp in length
11507: gap of unknown length
11607: contig of 1318 bp in length
12925: gap of unknown length
13025: contig of 1533 bp in length
14558: gap of unknown length
14657: contig of 1349 bp in length
14658: gap of unknown length
16007: contig of 1498 bp in length
16107: gap of unknown length
17604: contig of 1432 bp in length
17705: gap of unknown length

19137: gap of unknown length
19237: contig of 1139 bp in length
20375: gap of unknown length
20475: contig of 1162 bp in length
21637: gap of unknown length
21737: contig of 1264 bp in length
21738: gap of unknown length
23002: contig of 1349 bp in length
23102: gap of unknown length
24451: contig of 1724 bp in length
24551: gap of unknown length
26275: contig of 1110 bp in length
26375: gap of unknown length
27485: contig of 1623 bp in length
27585: gap of unknown length
29208: contig of 1689 bp in length
29308: gap of unknown length
30997: contig of 1133 bp in length
31097: gap of unknown length
32230: contig of 1874 bp in length
32330: gap of unknown length
34204: contig of 1390 bp in length
34304: gap of unknown length
35694: contig of 1182 bp in length
35794: gap of unknown length
36976: contig of 1190 bp in length
37076: gap of unknown length
38265: contig of 1124 bp in length
38365: gap of unknown length
39489: contig of 1036 bp in length
39590: gap of unknown length
39990: contig of 1266 bp in length
40626: gap of unknown length
40726: contig of 1266 bp in length
41992: gap of unknown length
42091: contig of 1037 bp in length
43128: gap of unknown length
43228: contig of 1281 bp in length
44510: gap of unknown length
44519: contig of 1486 bp in length
44610: gap of unknown length
46096: contig of 2076 bp in length
46195: gap of unknown length
48271: contig of 1907 bp in length
48372: gap of unknown length
50278: contig of 3169 bp in length
50379: gap of unknown length
53547: contig of 1740 bp in length
53647: gap of unknown length
53648: contig of 1234 bp in length
55388: gap of unknown length
55487: contig of 1234 bp in length
56721: gap of unknown length
56821: contig of 2210 bp in length
59031: gap of unknown length
59131: contig of 1844 bp in length
59132: gap of unknown length
60975: contig of 1918 bp in length
61075: gap of unknown length
62993: contig of 1339 bp in length
63094: gap of unknown length
64432: contig of 2087 bp in length
64533: gap of unknown length
66619: contig of 2075 bp in length
66719: gap of unknown length
68795: contig of 2300 bp in length
68895: gap of unknown length
71195: contig of 2375 bp in length
71295: gap of unknown length
73670: contig of 2334 bp in length
73770: gap of unknown length
76103: contig of 2664 bp in length
76203: gap of unknown length
78867: contig of 1202 bp in length
78968: gap of unknown length
80170: contig of 1676 bp in length
80270: gap of unknown length
81945: contig of 1676 bp in length
82045: gap of unknown length

Unpublished

25177 263335: contig of 1159 bp in length
26335 26425: size of unknown length

* 26336

```

26436 28199: contig of 1764 bp in length
28200 28299: gap of unknown length
28300 29444: contig of 1145 bp in length
29445 29545: gap of unknown length
29545 31447: contig of 1903 bp in length
31448 31547: gap of unknown length
31548 34038: contig of 2491 bp in length
34039 34139: gap of unknown length
34139 35911: contig of 1673 bp in length
35912 38254: contig of 2243 bp in length
38255 38355: gap of unknown length
38355 39887: contig of 1532 bp in length
39887 39986: gap of unknown length
41723 41822: contig of 1736 bp in length
41823 43529: contig of 1706 bp in length
43529 43628: gap of unknown length
43629 45291: contig of 1662 bp in length
45291 45391: gap of unknown length
45391 47399: contig of 2009 bp in length
47399 47500: gap of unknown length
47500 49019: contig of 1519 bp in length
49019 49118: gap of unknown length
49118 51076: contig of 1958 bp in length
51077 51177: gap of unknown length
51177 53555: contig of 2379 bp in length
53555 53656: gap of unknown length
53656 56041: contig of 2386 bp in length
56042 56141: gap of unknown length
56142 58493: contig of 2351 bp in length
58493 58592: gap of unknown length
58592 60409: contig of 1817 bp in length
60410 60510: gap of unknown length
60510 63422: contig of 2913 bp in length
63422 63522: gap of unknown length
63522 63780: contig of 2258 bp in length
63780 65881: gap of unknown length
65881 68421: contig of 2541 bp in length
68422 68521: gap of unknown length
68521 71408: contig of 2887 bp in length
71409 71508: gap of unknown length
71508 74335: contig of 2827 bp in length
74335 74436: gap of unknown length
74436 76581: contig of 2146 bp in length
76581 76682: gap of unknown length
76682 79513: contig of 2831 bp in length
79513 82209: gap of unknown length
82209 82309: contig of 2597 bp in length
82309 86599: gap of unknown length
86599 86600: contig of 4290 bp in length
86600 90718: gap of unknown length
90718 90819: contig of 4019 bp in length
90819 94041: gap of unknown length
94041 94140: contig of 3222 bp in length
94140 97476: gap of unknown length
97476 97577: contig of 3336 bp in length
97577 101462: gap of unknown length
101462 101562: contig of 3886 bp in length
101562 105925: gap of unknown length
105925 106025: contig of 4363 bp in length
106025 110657: gap of unknown length
110657 110758: contig of 4632 bp in length
110758 116260: gap of unknown length
116260 116360: contig of 5503 bp in length
116360 120661: gap of unknown length
120661 120761: contig of 4301 bp in length
120761 125559: gap of unknown length
125559 125659: contig of 4798 bp in length
125659 129795: gap of unknown length
129795 129895: contig of 4136 bp in length
129895 135463: gap of unknown length
135463 135463: contig of 5568 bp in length

```

Query Match 63.0%; Score 20.8; DB 2; Length 189192;
Best Local Similarity 78.1%; Pred. No. 1, le+02;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

DB 30724 CCTCATGTGTGCCAAGGCTCTGTGATGCGC 30693
 2 CCACAGAGTGGCAAGGCTCTGTGATGACCG 33
 11 111 111 111111111111 11

RESULT 7
LOCUS AC109903
DEFINITION AC109903
ACCESSION AC109903
VERSION AC109903.3 GI:21738286
KEYWORDS HTG, HTGS, PHASE1.
SOURCE Notway rat.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
AUTHORS 1 (bases 1 to 205020)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-ouman, F.R., Allen, C., Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayala, M., Banks, T., Barbieri, J., Benton, J., Blum, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brileva, M., Brown, M., Brown, N., Bryant, N.P., Buhay, C., Burch, P., Burdett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthett, K.J., Draper, H., Dugan-Rocha, S., Dublin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabriel, A., Gao, J., Garcia, A., Garner, F., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsl, F., Howard, S., Huber, J., Hult, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudh, S., Karlsson, E., Kelly, S., Khan, U., King, L., Kovach, J., Kover, C., Kratochvil, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S., Ogutu, M., Okwundu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Qulles, M., Ren, Y., Rives, M., Rojas, A., Rojoubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N., Slason, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Syatel, A., Taber, P., Tameisa, A., Tameisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, F., Telford, B., Thomas, N., Thomas, S., Uman, K., Vasquez, L., Vera, V., Villalón, D., Vihson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE Unpublished
JOURNAL Direct Submission
REFERENCE 2 (bases 1 to 205020)
AUTHORS Worley, K.C.
JOURNAL Direct Submission
REFERENCE Submitted (08-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 205020)

Db 197749 CCMCARTGTGCCAGCTTCGTGATGATGCG 197780

RESULT 8
 AC102955
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-32J3, *** SEQUENCING IN PROGRESS ***
 AC102955
 AC102955.3 GI:21723398
 HTG: HTG.PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 318696)
 Muny,D.M., Adams,C., Adlo-Oduola,B., All-oman,F.R., Allen,C.,
 Albrooks,S.L., Amaralung,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonin,D.,
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Bunay,C., Butch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,T., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogue,N., Holloway,C., Hollins,B.,
 Homs,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisse,D.,
 Lozodo,R.J., Lu,X., Lucier,A., Lucier,L., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
 Miner,G., Miner,Z., Mitchell,T., Monabhat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokwkw,S., Ogutu,M., Okwomou,G.,
 Orungu,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Severy,G.,
 Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Swalek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczky,R., Woden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G., and Gibbs,R.
 Direct Submission
 Unpublished
 JOURNAL 21 (bases 1 to 318696)
 TITLE
 REFERENCE
 AUTHORS
 JOURNAL
 Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 31 (bases 1 to 318696)
 Worley,K.C.
 Direct Submission
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 10, 2002 this sequence version replaced gi:179439342.
 COMMENT
 Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project information
 Center project name: GERN
 Center clone name: CH230-32J3
 Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye 1000 of reads
 Assembly program: Phrap version 0.990329
 Consensus quality: 163376 bases at least Q40
 Consensus quality: 175615 bases at least Q30
 Consensus quality: 184652 bases at least Q20
 NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 167 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 1
 1029: contig of 1029 bp in length
 1030
 1129: gap of unknown length
 1130
 2374: contig of 1245 bp in length
 2375
 2474: gap of unknown length
 2475
 3507: contig of 1033 bp in length
 3607: gap of unknown length
 3608
 4647: contig of 1040 bp in length
 4648
 4747: gap of unknown length
 4748
 5751: contig of 1004 bp in length
 5752
 5851: gap of unknown length
 5852
 7132: contig of 1261 bp in length
 7133
 7233: gap of unknown length
 8348: contig of 1116 bp in length
 8349
 8449: gap of unknown length
 8449
 9634
 9733: contig of 1185 bp in length
 9734
 10735: gap of unknown length
 10736
 10835: contig of 1002 bp in length
 10836
 12255: gap of unknown length
 12256
 12355: gap of unknown length
 12356
 13491: contig of 1136 bp in length
 13492
 13591: gap of unknown length
 13592
 14685: contig of 1094 bp in length
 14686
 14785: gap of unknown length
 14786
 16277: contig of 1492 bp in length
 16278
 16377: gap of unknown length
 16378
 17828: contig of 1451 bp in length
 17829
 17928: gap of unknown length
 17929
 19181: contig of 1253 bp in length
 19182
 19282: gap of unknown length
 19283
 20409: contig of 1128 bp in length
 20410
 20509: gap of unknown length
 20510
 21539: contig of 1030 bp in length
 21540
 21639: gap of unknown length
 21640
 22995: contig of 1336 bp in length
 22996
 23095: gap of unknown length
 23096
 24553: contig of 1458 bp in length
 24554
 24653: gap of unknown length
 24654
 25699: contig of 1046 bp in length
 25700
 25799: gap of unknown length
 25800
 26813: contig of 1014 bp in length
 26814
 26913: gap of unknown length
 26914
 28457: contig of 1544 bp in length
 28458
 28557: gap of unknown length
 28558
 30081: contig of 1524 bp in length
 30082
 30181: gap of unknown length
 30182
 31453: contig of 1272 bp in length
 31454
 31553: gap of unknown length

```

* 31554 32635: contig of 1082 bp in length
* 32636 32735: gap of unknown length
* 32736 33742: contig of 1007 bp in length
* 33743 33842: gap of unknown length
* 33843 35401: contig of 1559 bp in length
* 35402 35502: gap of unknown length
* 35502 36514: contig of 1013 bp in length
* 36515 37748: contig of 1134 bp in length
* 37749 37849: gap of unknown length
* 37849 39053: contig of 1205 bp in length
* 39054 39153: gap of unknown length
* 39154 40437: contig of 1284 bp in length
* 40438 40537: gap of unknown length
* 40538 41998: contig of 1461 bp in length
* 41999 42098: gap of unknown length
* 42099 43458: contig of 1360 bp in length
* 43459 43558: gap of unknown length
* 43559 44573: contig of 1015 bp in length
* 44574 44673: gap of unknown length
* 44674 45731: contig of 1058 bp in length
* 45732 45831: gap of unknown length
* 45832 47490: contig of 1659 bp in length
* 47491 47591: gap of unknown length
* 47591 48619: contig of 1029 bp in length
* 48620 48719: gap of unknown length
* 48720 49748: contig of 1029 bp in length
* 49749 49848: gap of unknown length
* 49849 51127: contig of 1279 bp in length
* 51128 51227: gap of unknown length
* 51228 52354: contig of 1127 bp in length
* 52355 52454: gap of unknown length
* 52455 53733: contig of 1279 bp in length
* 53734 53833: gap of unknown length
* 53834 54939: contig of 1106 bp in length
* 54940 55039: gap of unknown length
* 55040 56467: contig of 1428 bp in length
* 56468 57628: gap of unknown length
* 57629 57728: gap of unknown length
* 57729 58837: contig of 1109 bp in length
* 58838 58937: gap of unknown length
* 58938 59946: contig of 1009 bp in length
* 59947 60046: gap of unknown length
* 60047 61049: contig of 1003 bp in length
* 61050 62708: gap of unknown length
* 62709 62808: gap of unknown length
* 62809 64277: contig of 1469 bp in length
* 64278 64377: gap of unknown length
* 64378 66150: contig of 1773 bp in length
* 66151 66250: gap of unknown length
* 66251 67433: contig of 1183 bp in length
* 67434 67533: gap of unknown length
* 67534 69376: contig of 1843 bp in length
* 69377 70831: gap of unknown length
* 70832 70931: contig of 1355 bp in length
* 70932 72325: gap of unknown length
* 72326 73325: contig of 1394 bp in length

```

Query Match 63.0%; Score 20.8; DB 2; Length 318696;
 Best Local Similarity 78.1%; Pred. No. 1.1e+02;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 CCACATGATGCGCAAGCGTCTGTGATACCG 33
 Db 298103 CCTCATGTGTCGCAAGCGTCTGTGATACCG 298134

RESULT 9
 DB9010 3079 bp DNA 1linear PLN 12-MAY-1999
 LOCUS
 DEFINITION Aspergillus oryzae DNA for BRLA, complete cds.

ACCESSION DB9010
 VERSION DB9010.1 GI:1694664
 KEYWORDS BRLA.
 SOURCE Aspergillus oryzae DNA.
 ORGANISM Aspergillus oryzae
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 REFERENCE
 AUTHORS Yamada, O., Lee, R. R., Gomi, K. and Iimura, Y.
 TITLE Cloning and functional analysis of the Aspergillus oryzae
 conidiation regulator gene brla by its disruption and misscheduled
 expression
 JOURNAL J. Biosci. Bioeng. 87, 424-429 (1999)
 REFERENCE
 AUTHORS Yamada, O.
 TITLE Direct Submission
 JOURNAL Submitted (12-NOV-1996) Osamu Yamada, National Research Institute
 of Brewing, Microbiology Division; 3-7-1 Kagamiyama,
 Hiasahi-Hiroshima, Hiroshima 739, Japan
 (E-mail: Yamada.Oenr@b.jp, Tel: 81-824-20-0824,
 Fax: 81-824-20-0808)

FEATURES
 source
 1..3079
 Location/Qualifiers
 /organism="Aspergillus oryzae"
 /db_xref="taxon:5062"
 1449..2714
 /gene="brla"
 1449..2714
 /gene="brla"
 /codon_start=1
 /product="BRLA"
 /protein_id="BAA13744.1"
 /db_xref="GI:1694665"
 /translation="MRTQNNLTVEVDCHSLGNSNCPSTSSFSFMDSEPTPTPTISYQ
 GSLASPGMODAGSYGHAVERHTGATPMSAFRLAGTNSNMGMSCGAMBAOERMPE
 PDFLSAYDNDYHEWMLSDGKRYETGSHPLTPOCPMYRSNVPRAATLPEA
 ATNCLASRSIFNHAERVPDSMGNMAMPVIOAVESIAPORTIASOVGVPYPPSYSE
 PFTSIQFKHTSPPTPLRSCGISGASGPDPTISLGGAAATLDFDFOOSPFRGLNR
 LORPDSRMTRKSSRONMSIENPSTIKOVQFPCKEGCGRRGRDHLKRHHKSL
 KERHVCNVECCERAFSRSDMLNHYTHSKRGGRNRYATVLDSESFYDPPDRGOL
 TPDSLPTRGSLTDDMPNRSRYSDGLDD"

BASE COUNT 728 a 858 c 587 g 906 t
 ORIGIN
 Query Match 61.8%; Score 20.4; DB 8; Length 3079;
 Best Local Similarity 80.0%; Pred. No. 1.7e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 OY 3 CACATGATGCGCAAGCGTCTGTGATACCG 32
 Db 1394 CACCTGGGTGCGCAAGCGCGGTCTATACCG 1623

RESULT 10
 LOCUS DMCL118B3/c 24263 bp DNA 1linear INV 27-APR-1999
 DEFINITION Drosophila melanogaster cosmid clone 118B3.
 ACCESSION AL031582
 VERSION AL031582.1 GI:4176433
 KEYWORDS
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE
 AUTHORS Philippe, Valenti., Catherine, Salles., Lorna, Campbell. and
 David, Glover.
 TITLE Sequencing the distal X chromosome of Drosophila melanogaster
 JOURNAL Unpublished
 DEPARTMENT Department of Anatomy and Physiology, Medical Sciences Institute,
 University of Dundee, Dundee DD1 4HN, U.K.
 REFERENCE 2 (bases 1 to 24263)

DEFINITION Homo sapiens chromosome 11 clone RP5-1135K18 map 11, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC087531 GI:13431031
VERSION AC087531.3 GI:13431031
KEYWORDS HTG: HTGS-PHASEO.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 40396)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE 1 Homo sapiens chromosome 11, clone RP5-1135K18
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 40396)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Batra,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Fitzhugh,W., Gage,D., Galagan,J.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Lakoque,K., Lamazares,R., Landers,T.,
Lenczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Marquis,N., Mathews,C., McCarthy,M., McEwan,P., McKernan,K.,
Mcneesters,R., Meldrim,J., Menus,L., Mihova,T., Mlenka,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Riedack,M., Riley,R., Rise,C., Kogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Soungas,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 22, 2001 this sequence version replaced gi:13184089.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Genome Center
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: 112143
Center clone name: 1135_K_18

* NOTE: This record contains 62 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1. 570: contig of 570 bp in length
* 571 670: gap of 100 bp
* 671 1229: contig of 559 bp in length
* 1230 1329: gap of 100 bp
* 1330 1887: contig of 558 bp in length
* 1888 1987: gap of 100 bp
* 1988 2527: contig of 540 bp in length
* 2528 2627: gap of 100 bp
* 2628 3176: contig of 549 bp in length
* 3177 3276: gap of 100 bp
* 3277 3829: contig of 553 bp in length

3830 3929: gap of 100 bp
3930 4492: contig of 563 bp in length
4493 4592: gap of 100 bp
4593 5148: contig of 556 bp in length
5149 5248: gap of 100 bp
5249 5781: contig of 533 bp in length
5782 5881: gap of 100 bp
5882 6442: contig of 561 bp in length
6443 6542: gap of 100 bp
6543 7096: contig of 554 bp in length
7097 7196: gap of 100 bp
7197 7743: contig of 547 bp in length
7744 7843: gap of 100 bp
7844 8394: contig of 551 bp in length
8395 8494: gap of 100 bp
8495 9054: contig of 560 bp in length
9055 9154: gap of 100 bp
9155 9693: contig of 539 bp in length
9694 9793: gap of 100 bp
9794 10345: contig of 557 bp in length
10346 10445: gap of 100 bp
10446 10998: contig of 553 bp in length
10999 11098: gap of 100 bp
11099 11634: contig of 536 bp in length
11635 11734: gap of 100 bp
11735 12296: contig of 562 bp in length
12297 12396: gap of 100 bp
12397 12971: contig of 575 bp in length
12972 13071: gap of 100 bp
13072 13637: contig of 566 bp in length
13638 13737: gap of 100 bp
13738 14298: contig of 561 bp in length
14299 14398: gap of 100 bp
14399 14957: contig of 555 bp in length
14958 15057: gap of 100 bp
15058 15588: contig of 531 bp in length
15589 15688: gap of 100 bp
15689 16245: contig of 557 bp in length
16246 16345: gap of 100 bp
16346 16899: contig of 554 bp in length
16899 16999: gap of 100 bp
16900 17554: contig of 555 bp in length
17555 17654: gap of 100 bp
17655 18209: contig of 555 bp in length
18210 18309: gap of 100 bp
18310 18867: contig of 556 bp in length
18868 18967: gap of 100 bp
18968 19505: contig of 538 bp in length
19506 19605: gap of 100 bp
19606 20169: contig of 564 bp in length
20170 20269: gap of 100 bp
20270 20831: contig of 562 bp in length
20832 20931: gap of 100 bp
20932 21479: contig of 548 bp in length
21480 21579: gap of 100 bp
21580 22141: contig of 562 bp in length
22142 22241: gap of 100 bp
22242 22794: contig of 553 bp in length
22795 22894: gap of 100 bp
22895 23445: contig of 551 bp in length
23446 23545: gap of 100 bp
23546 24074: contig of 529 bp in length
24075 24174: gap of 100 bp
24175 24723: contig of 549 bp in length
24724 24823: gap of 100 bp
24824 25374: contig of 551 bp in length
25375 25474: gap of 100 bp
25475 26022: contig of 548 bp in length
26023 26122: gap of 100 bp
26123 26683: contig of 561 bp in length
26684 26783: gap of 100 bp
26784 27303: contig of 520 bp in length
27304 27403: gap of 100 bp

```

27404 27958: contig of 555 bp in length
27959 28058: gap of 100 bp
28059 28614: contig of 556 bp in length
28615 28714: gap of 100 bp
28715 29270: contig of 556 bp in length
29271 29370: gap of 100 bp
29371 29928: contig of 558 bp in length
29929 30028: gap of 100 bp
30029 30580: contig of 552 bp in length
30581 30680: gap of 100 bp
31229 31229: contig of 549 bp in length
31230 31329: gap of 100 bp
31330 31682: contig of 553 bp in length
31683 31982: gap of 100 bp
31983 32538: contig of 556 bp in length
32539 32638: gap of 100 bp
32639 33187: contig of 549 bp in length
33188 33287: gap of 100 bp
33288 33837: contig of 550 bp in length
33838 33937: gap of 100 bp
33938 34484: contig of 547 bp in length
34485 34584: gap of 100 bp
34585 35145: contig of 561 bp in length
35146 35245: gap of 100 bp
35246 35799: contig of 554 bp in length
35800 35899: gap of 100 bp
35900 36445: contig of 546 bp in length
36446 36545: gap of 100 bp
36546 37103: contig of 558 bp in length
37104 37203: gap of 100 bp
37204 37760: contig of 557 bp in length
37761 37860: gap of 100 bp
37861 38424: contig of 564 bp in length
38425 38524: gap of 100 bp
38525 39071: contig of 547 bp in length
39072 39171: gap of 100 bp
39172 39735: contig of 564 bp in length
39736 39835: gap of 100 bp
39836 40396: contig of 561 bp in length.
Location/Qualifiers
1. 40396
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone_id="RP5-1135K18"
/clone_lib="RPCI Human PAC library 5"
9313 a 7763 c 7710 g 9183 t 6427 others

Query Match
Best Local Similarity 75.8%; Score 20.2; DB 2; Length 40396;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 GCCACATGAGTGGCAAGCGCTGTGGTATCCG 33
Db 6231 GCCACAGAGGAGGCGAGGAGCTGTGGTATCCG 6199

RESULT 12
AC020129 102227 bp DNA linear HTG 03-JAN-2000
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
ACCESSION AC020129
VERSION AC020129.1 GI:6664768
KEYWORDS HTG; PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephyridroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 102227)

```

```

AUTHORS Adams M. and Venter J.C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10212145 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..102227
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 27684 a 23342 c 23493 g 27708 t
ORIGIN
Query Match
Best Local Similarity 75.8%; Score 20.2; DB 2; Length 102227;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 GCCACATGAGTGGCAAGCGCTGTGGTATCCG 33
Db 91773 GTCGAGAGATGTGACGCGCTGTGGGATCCG 91805

RESULT 13
AL133294/c 106265 bp DNA linear PRI 27-SEP-2000
LOCUS Human DNA sequence from clone RP5-1135K18 on chromosome 11p12-14.1
ACCESSION AL133294
VERSION AL133294
KEYWORDS Contains GSSs, complete sequence.
SOURCE AL133294-10 GI:9650513
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 106265)
AUTHORS Moore M.
JOURNAL Direct Submission
COMMENT Submitted (15-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Aug 1, 2000 this sequence version replaced gi:9501134.
During difference assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
corresponding to the overlapping clone name. Note that the
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information
on the WormPep database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormPep
RP5-1135K18 is
institute by the group of Pieter de Jong. For further details see
http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RP5-1135K18 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP5-1135K18 is at 1 in this sequence.
The true left end of clone RP5-1169J3 is at 106166 in this
sequence.
Location/Qualifiers

```

```
source
1. 106265
/organism="Homo sapiens"
/dbxref="taxon:9606"
/chromosome="11"
/map="p12-14.1"
/clone="RP5-1135K1.8"
/clone.lib="RPC1-5"
1279. 1591
repeat_region
/note="AluX repeat: matches 1. 311 of consensus"
1592. 1675
repeat_region
/note="42 copies 2 mer tt 71% conserved"
1701. 1895
repeat_region
/note="L1ME3 repeat: matches 5927. 6123 of consensus"
1902. 2049
repeat_region
/note="MIR repeat: matches 92. 249 of consensus"
2908. 3094
repeat_region
/note="L2 repeat: matches 2426. 2624 of consensus"
3089. 3363
misc_feature
/note="match: GSS: Em:B01382"
complement(3349. 3742)
/note="match: GSS: Em:AQ019578"
4577. 4790
repeat_region
/note="MIR repeat: matches 48. 261 of consensus"
6479. 6734
repeat_region
/note="MIR repeat: matches 6. 262 of consensus"
7746. 7969
repeat_region
/note="L1MEC repeat: matches 2177. 2401 of consensus"
7970. 8277
repeat_region
/note="AluX repeat: matches 1. 303 of consensus"
8278. 8419
repeat_region
/note="L1MEC repeat: matches 2377. 2177 of consensus"
8420. 8724
repeat_region
/note="AluX repeat: matches 1. 298 of consensus"
8725. 8846
repeat_region
/note="L1MEC repeat: matches 2257. 2378 of consensus"
9041. 9084
repeat_region
/note="22 copies 2 mer tt 81% conserved"
9087. 9396
repeat_region
/note="AluX repeat: matches 1. 312 of consensus"
9479. 9676
repeat_region
/note="MIR repeat: matches 49. 262 of consensus"
9967. 10097
repeat_region
/note="Aluub repeat: matches 3. 98 of consensus"
complement(10129. 10648)
/note="match: GSS: Em:AQ071611"
complement(10161. 10661)
/note="match: GSS: Em:AQ353682"
complement(10214. 10681)
/note="match: GSS: Em:AQ386199"
10796. 10853
repeat_region
/note="MIR repeat: matches 149. 207 of consensus"
10855. 11158
repeat_region
/note="AluSq repeat: matches 1. 308 of consensus"
11251. 11515
repeat_region
/note="L2 repeat: matches 2451. 2749 of consensus"
11648. 11851
repeat_region
/note="L1MC5 repeat: matches 7705. 7917 of consensus"
11859. 12150
repeat_region
/note="AluX repeat: matches 1. 295 of consensus"
12175. 12366
repeat_region
/note="Aluub repeat: matches 89. 273 of consensus"
12472. 12743
repeat_region
/note="AluX repeat: matches 26. 301 of consensus"
complement(13148. 13905)
/note="match: GSS: Em:B05804"
complement(13300. 13785)
/note="match: GSS: Em:AQ32970"
13309. 13539
repeat_region
/note="MIR repeat: matches 262. 478 of consensus"
complement(13345. 13683)
/note="match: GSS: Em:AQ091582"
complement(13384. 13771)
/note="match: GSS: Em:AQ005856"

misc_feature
complement(13441. 13785)
/note="match: GSS: Em:B50430"
complement(13461. 13785)
/note="match: GSS: Em:B78931 Em:B88270"
complement(13829. 14267)
/note="match: GSS: Em:AQ228611"
13858. 14164
repeat_region
/note="AluSp repeat: matches 1. 308 of consensus"
complement(14001. 14450)
/note="match: GSS: Em:AQ428027"
complement(14063. 14269)
/note="match: GSS: Em:B53280"
15018. 15097
repeat_region
/note="AluS repeat: matches 2. 82 of consensus"
15098. 15402
repeat_region
/note="AluX repeat: matches 1. 306 of consensus"
15415. 15540
repeat_region
/note="MIR repeat: matches 13. 144 of consensus"
15775. 16192
misc_feature
/note="match: GSS: Em:AQ084215"
16628. 16795
repeat_region
/note="L1PA16 repeat: matches 5823. 5965 of consensus"
16796. 17090
repeat_region
/note="AluX repeat: matches 1. 288 of consensus"
17091. 17263
repeat_region
/note="L1PA16 repeat: matches 5965. 6146 of consensus"
17286. 17574
repeat_region
/note="Aluub repeat: matches 1. 300 of consensus"
20176. 20349
repeat_region
/note="MER45 repeat: matches 2. 178 of consensus"
20396. 20571
repeat_region
/note="MIR repeat: matches 64. 244 of consensus"
20607. 20731
repeat_region
/note="L2 repeat: matches 2627. 2749 of consensus"
21239. 21369
repeat_region
/note="MIR repeat: matches 116. 262 of consensus"
22075. 22197
repeat_region
/note="MIR repeat: matches 19. 148 of consensus"
22199. 22357
repeat_region
/note="Aluub repeat: matches 135. 293 of consensus"
22895. 23206
repeat_region
/note="AluX repeat: matches 1. 312 of consensus"
24362. 24506
repeat_region
/note="L2 repeat: matches 2308. 2439 of consensus"
24840. 24877
repeat_region
/note="L2 repeat: matches 1656. 1693 of consensus"
25110. 25504
repeat_region
/note="MER69 repeat: matches 1995. 2454 of consensus"
25503. 25653
repeat_region
/note="MER69 repeat: matches 419. 568 of consensus"
25738. 26004
repeat_region
/note="MER69 repeat: matches 1. 303 of consensus"
26167. 26477
repeat_region
/note="AluSq repeat: matches 1. 313 of consensus"
26836. 27008
repeat_region
/note="MIR repeat: matches 89. 262 of consensus"
27070. 27363
repeat_region
/note="AluSq repeat: matches 1. 295 of consensus"
28049. 28082
repeat_region
/note="L7 copies 2 mer tt 94% conserved"
28140. 28376
misc_feature
/note="match: GSS: Em:AQ079942"
28309. 28488
repeat_region
/note="FRAM repeat: matches 0. 176 of consensus"
28521. 28828
repeat_region
/note="AluX repeat: matches 3. 311 of consensus"
28865. 29257
misc_feature
/note="match: GSS: Em:AQ061107"
29126. 29270
repeat_region
/note="L1MC4 repeat: matches 7676. 7815 of consensus"
29271. 29366
repeat_region
/note="L1MC4 repeat: matches 7877. 7975 of consensus"
30243. 30349
repeat_region
```

	Query March	61.2%	Score 20.2;	DB 9;	Length 106265;
	Best Local Similarity	75.8%	Pred No. 2; e=0.2;		
	Matches	25;	Conservative	0;	Mismatches 8;
					Indels 0;
					Gaps 0;
Oy	1 GCCACATGAGTGCAGAGCGCTGGTGATTACCG	33			
Dh	92630 GCCACAGGAGGGCGCAGGAGGTGGTAGTCACG	92598			

RESULT 14	AC104602/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	AC104602		161668 bp DNA linear INV 06-APR-2002	Drosophila melanogaster X BAC R98-1666 (Roswell Park Cancer Institute Drosophila BAC library) complete sequence.	AC104602.3	GI:18030109	HTG.	Drosophila melanogaster.
								Drosophila melanogaster.

REFERENCE
AUTHORS
Muzny, D., Scherer, S., Adams, M. D., Tringali, J. J., and White, J. 1993. The *Drosophila* genome project: bases 1 to 161668). *Drosophila*.

TITLE
Direct Submission
Davenport, L.B., Dietz, S.M., Ding, Y., Dodson, K., Doup, L.E.,
Draper, H., Emery-Cohen, A., Ferreira, S., Gary, N.D.S., Houck, J.,
Hostin, D., Howland, A., Idegawa, C., Jajall, M., Kovar, K.,
Liu, M., Matted, B., McInosh, T.C., Morgan, M., Moy, M., Murphy, B.,
Nelson, K.A., Ndassa, Y., Nguyen, N., Perez, L., Pittman, G.S., Puri, V.,
Schneider, F., Shen, H., Strong, R., Tector, C., Wang, Q., Williams, S.M.,
Xiang, J., Zaveril, J.S., Zhou, J., Zorrilla, S., Smith, H.O.,
Weier, D., Weinstein, G., Gibbs, R. and Venter, J.C.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

Direct Submission
 Unpublished
 2 (bases 1 to 161668)
 Worley, K.C.
 Direct Submission
 Submitted (14-DEC-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza Houston, TX 77030, USA
 3 (bases 1 to 161668)
 Worley, K.C.

31 (Pages 1 to 16168)

REFERENCE
AUTHORS

Worley K.C., Adams C.,
Alshrooks, S.L., Amaralunga, H.C., Ali- Osman, F.R., Allen, C.,
Benton, J., Blimage, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowe, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Butcher, P., Burrell, C., Burrell, K.L., Byrd, N.C., Carion, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Garroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Deann, A.L., Ding, Y., Din, H. H., Douthwaite, K.J., Diaper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Franz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Gatz, N., Giller, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C.,
Hollins, B., Homsl, F., Howard, S., Huber, J., Huliy, S., Hume, J.,

TITLE Direct Submission
JOURNAL Submitted (01-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (pages 1 to 161668)
AUTHORS Worley, R.C., Adams, C., Adio-Ogbonla, P., et al.

Alsbrooks, S. L., Amaratingue, H. C., Ara, J. R., Banks, T., Barbara, J., Benton, J., Blinige, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowles, S., Brivea, M., Brown, E., Brown, M., Bryant, N. P., Bunhy, C., Burch, P., Butrell, C., Butrell, K. L., Byrd, N. C., Carron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowhry, I., Christopoulos, C., Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R., Davila, M. L., Davis, C., Denny, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Dreaper, H., Dugan-Hochs, S., Durbin, K. J., Earnhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havtack, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Joliver, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kora, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L. J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Matthez, E., Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabhat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nicholson, E., Nwokweto, S., Ogun, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Peters, L., Peters, L., Pickens, R., Piment, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojoudokan, I., Rolfe, M., Ruiz, S., Severy, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodegriem, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swalek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Vera, V., Villalon, D., Thomas, S., Thomas, S., Usmani, K., Vasquez, L., Verra, V., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., and Gibbs, R.

TITLE Direct Submission
JOURNAL Submitted (06-Apr-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jan 1, 2002 this sequence version replaced gi:17976408.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc.help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:

STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT REPORT:

Location/Qualifiers

1. 161668

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="X"

/clone="RP98-16G6"

49371. .51271

/function="unresolved tandem repeat"

50505. .50507

/function="low quality region"

50533. .50534

/function="low quality region"

50547

/note="G/C"

/function="polymorphic site"

50552

/note="C/T"

/function="polymorphic site"

50579

/note="G/C"

/function="polymorphic site"

50588

/note="C/T"

/function="polymorphic site"

BASE COUNT 44985 a 36812 c 36365 g 43506 t

ORIGIN

Query Match 61.2%; Score 20.2; DB 3; Length 161668;

Best Local Similarity 75.8%; Pred. No. 2.2e+02; Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCCACATGAGTGCAGGCGCTGCTGATACCG 33

DB 85717 GTCGAGGAGTGTACGCGCTCTGGGATCCG 85685

RESULT 15

AC087522/c

LOCUS

DEFINITION

AC087522

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

1 (bases 1 to 179752)

AC087522 179752 bp DNA linear HTG 27-MAY-2001
Homo sapiens chromosome 11 clone RP11-567H10 map 11, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
AC087522
AC087522.5 GI:14210552
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Homo sapiens.
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 179752)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 11, clone RP11-567H10
Unpublished
2 (bases 1 to 179752)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A.,
Cammarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., Lacroque, K., Lamazares, R., Landers, T.,
Lenczky, J., Levine, R., Liu, G., Maclean, C., MacDonald, P.,
Marguis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheters, R., Meldrum, J., Menus, L., Mihova, T., Mlenka, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
Sougnaz, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.-J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 27, 2001 this sequence version replaced g1:14196392.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
Project Information
Center project name: 567_H_10
Center clone name: 567_H_10
Sequencing Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 177062 bases at least Q40
Consensus quality: 177860 bases at least Q20
Consensus quality: 178367 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 179152; sum-of-contigs
Quality coverage: 15.6 in Q20 bases; agarose-fp
Quality coverage: 15.3 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 7 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 664: contig of 664 bp in length

```

* 665 764: gap of 100 bp
* 765 1403: contig of 639 bp in length
* 1404 1503: gap of 100 bp
* 1504 98418: contig of 96915 bp in length
* 98419 98518: gap of 100 bp
* 98519 109646: contig of 11128 bp in length
* 109647 109746: gap of 100 bp
* 109747 137562: contig of 27816 bp in length
* 137563 137662: gap of 100 bp
* 137663 176971: contig of 39309 bp in length
* 176972 177071: gap of 100 bp
* 177072 179752: contig of 2681 bp in length.
Location/Qualifiers
1. 179752
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone_11b="RPC1-11 Human Male BAC"
1. 664
/misc_feature
/note="assembly-fragment"
765. 1403
/misc_feature
/note="assembly-fragment"
1504. 98418
/misc_feature
/note="assembly-fragment"
98519. 109646
/misc_feature
/note="assembly-fragment"
109747. 137562
/misc_feature
/note="assembly-fragment"
137663. 176971
/misc_feature
/note="assembly-fragment"
177072. 179752
/clone_end:17
vector_side:right"

```

```

BASE COUNT 46238 a 43601 c 43184 g 46106 t 623 others
ORIGIN
Query Match 61.2% Score 20.2; DB 2; Length 179752;
Best Local Similarity 75.8% Pred No. 2.2e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

OY 1 GCCACATGATGCGCAGGCGTGTGTGATACCG 33
DB 65594 GCCACAGGAGGGGCGAGGAGGTGTGATCCAG 65562

```

Search completed: July 10, 2003, 19:20:20
 Job time : 294.951 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 69.6337 Seconds
(Without alignments)
1067.241 Million cell updates/sec

Title: US-10-087-631B-5

Perfect score: 33
Sequence: 1 gccacatgagtgagcaaggcgtctgtgataccg 33

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.GeneSeq_101002:*

- 1: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1980.DAT:*
- 2: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1981.DAT:*
- 3: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1982.DAT:*
- 4: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1983.DAT:*
- 5: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1984.DAT:*
- 6: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1985.DAT:*
- 7: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1986.DAT:*
- 8: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1987.DAT:*
- 9: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1988.DAT:*
- 10: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1989.DAT:*
- 11: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1990.DAT:*
- 12: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1991.DAT:*
- 13: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1992.DAT:*
- 14: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1993.DAT:*
- 15: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1994.DAT:*
- 16: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1995.DAT:*
- 17: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1996.DAT:*
- 18: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1997.DAT:*
- 19: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1998.DAT:*
- 20: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1999.DAT:*
- 21: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA2000.DAT:*
- 22: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA2001A.DAT:*
- 23: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA2001B.DAT:*
- 24: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.8	60.0	771	22	AA6584
2	19.4	58.8	415	21	AAA67547
3	19.2	58.2	634	21	AAFL5168
4	19.2	58.2	906	21	AAFO7727
5	19.2	58.2	1284	23	AA565203
6	19.2	58.2	5660	20	AA222700
7	19	57.6	379	24	ABN18387
8	18.8	57.0	841	23	AA583486
9	18.6	56.4	472	24	ABN73219

C	10	18.6	56.4	973	21	AA633442	Arabidopsis thaliana
C	11	18.6	56.4	1122	23	AA583436	DNA encoding novel
C	12	18.6	56.4	249487	24	ABN85733	mouse genomic reg1
C	13	18.4	55.8	1131	21	AAA96719	Polynucleotide iso
C	14	18.4	55.8	3425	22	AA593747	Human cDNA encoding
C	15	18.2	55.2	518	21	AA248806	Rice inositol 1,3,
C	16	18.2	55.2	755	22	ABAL3910	Human nervous syst
C	17	18.2	55.2	853	22	AA106372	Human reproductive
C	18	18.2	55.2	1740	18	AA742749	RHDV, capsid gene.
C	19	18.2	55.2	4372	22	ABA20664	Human nervous syst
C	20	18.2	55.2	7785	20	AAV81446	Pig p105 zona pell
C	21	18	54.5	350	23	AA582637	DNA encoding novel
C	22	18	54.5	351	15	AAO63907	Heavy chain variab
C	23	18	54.5	446	23	AA570318	DNA encoding novel
C	24	18	54.5	500	22	AAK77964	Human immune/haema
C	25	18	54.5	593	22	ABA60417	Human foetal liver
C	26	18	54.5	593	22	AAK08695	Human brain expres
C	27	18	54.5	593	22	AAK34582	Human bone marrow
C	28	18	54.5	593	22	AA140301	Probe #8987 used t
C	29	18	54.5	593	24	AB509213	Human genome-deriv
C	30	18	54.5	654	24	ABO60434	Human colon cancer
C	31	18	54.5	864	23	AA570317	DNA encoding novel
C	32	18	54.5	1014	23	AA564291	DNA encoding novel
C	33	18	54.5	1014	23	AA571765	DNA encoding novel
C	34	18	54.5	1014	23	AA588225	DNA encoding novel
C	35	18	54.5	1100	22	AA531252	Human immune respo
C	36	18	54.5	1140	23	AA581729	DNA encoding novel
C	37	18	54.5	1322	23	AA581757	DNA encoding novel
C	38	18	54.5	2187	23	AA580450	DNA encoding novel
C	39	18	54.5	2541	23	AA588228	DNA encoding novel
C	40	18	54.5	2658	23	AA581745	DNA encoding novel
C	41	18	54.5	2660	23	AA582684	DNA encoding novel
C	42	18	54.5	2801	23	AA578375	DNA encoding novel
C	43	18	54.5	3663	23	ABL23920	Drosophila melanog
C	44	18	54.5	6261	23	ABL03774	Drosophila melanog
C	45	18	54.5	36434	24	AA631198	Human WKL-1 genom

ALIGNMENTS

RESULT 1	AA6584	standard; DNA; 771 BP.
ID	AA6584	
AC	AA6584	
DT	02-APR-2001	(first entry)
DE	DNA encoding a fusion of a single chain antibody and streptavidin.	
KW	Streptavidin; tumour cell; cancer; adenocarcinoma;	
KW	hemtological malignancy; ss.	
OS	Synthetic.	
OS	Streptomyces avidinii.	
OS	Unidentified.	
PN	W020075333-A1.	
PD	14-DEC-2000.	
PF	05-JUN-2000; 2000MO-US15595.	
PR	07-JUN-1999; 99US-0137900.	
PR	03-DEC-1999; 99US-0168976.	
PA	(NEOR-) NEORX CORP.	
PI	Goshorn SC, Graves SS, Schultz JE, Lin Y, Sanderson JA, Reno JM;	
DR	WPI; 2001-091213/10.	

PT New vector constructs for expressing genomic streptavidin fusion
 PT proteins which are useful for targeting tumour cells associated with
 PT cancer, e.g. adenocarcinomas -

PS Example 5; Page 93; 100bp; English.

CC The present sequence encodes a fusion of an anti-CEA single chain
 CC antibody and streptavidin. The fusion protein is expressed using
 CC vectors of the invention. The specification describes vector constructs
 CC for expressing streptavidin fusion proteins. The vector comprises a
 CC nucleic acid encoding streptavidin or its functional variant operatively
 CC linked to a promoter, and a cloning site for insertion of a second
 CC nucleic acid sequence encoding a polypeptide to be fused with
 CC streptavidin. Interposed between the promoter and the first nucleic
 CC acid sequence. Alternatively, the vector construct comprises a nucleic
 CC acid, operatively linked to a promoter, encoding a polypeptide to be
 CC fused with streptavidin, and a cloning site for insertion of a second
 CC nucleic acid encoding at least 129 amino acids of streptavidin or its
 CC functional variant. The fusion proteins are useful for targeting tumour
 CC cells, particularly tumour cells associated with cancer.
 CC e.g. adenocarcinomas or hematological malignancies. The vector construct
 CC is useful for expressing of streptavidin fusion proteins. In particular,
 CC these are useful as tools for medical diagnostics and therapeutic
 CC purposes, e.g. for detecting the presence or absence of, or treating, a
 CC target site within a mammalian host.

SQ Sequence 771 BP; 172 A; 180 C; 230 G; 189 T; 0 other;

Query Match Best Local Similarity 60.0%; Score 19.8; DB 22; Length 771;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 CCACATGATGTCGCAAGCGCTGCGTATACC 32
 DB 54 CAAGATCACTGCTGCAAGCGCTTGGTATACC 84

RESULT 2

ID AAA67547/c standard; DNA; 415 BP.

AC AAA67547;

DT 31-OCT-2000 (first entry)

DE Pinus radiata sucrose phosphatase synthase DNA SEQ ID NO:548.

KW Eucalyptus grandis; pinus radiata; Monterey pine; modification;
 KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
 KW transgenic plant; ds.

OS Pinus radiata.

PN WO200022092-A2.

PD 20-APR-2000.

PF 08-OCT-1999; 99WO-NZ00169.

PR 13-OCT-1998; 98US-0170862.

PR 11-AUG-1999; 99US-0148426.

PA (GENE-) GENESIS RES & DEV CORP LTD.

XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Bloksberg LN;

DR WPI; 2000-339328/29.

PT New genes encoding proteins involved in a plant polysaccharide
 PT biosynthetic pathway, useful for modulating or altering the
 PT polysaccharide content, composition or structure of the plant

PS Claim 1; Page 212-213; 301bp; English.

CC The present invention describes isolated polynucleotides (PN) comprising
 CC a sequence selected from one of 835 nucleotide sequences given in
 CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
 CC an expectation (E) value of 0.01 or less compared to the 835 sequences,
 CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
 CC 835 sequences or sequences that are degenerately equivalent or allelic
 CC to the 835 sequences. The polynucleotides are used to modify the
 CC activity of a polypeptide involved in a polysaccharide biosynthetic
 CC pathway in the plant. They are especially used to modulate or alter the
 CC polysaccharide content, composition or structure of the plant. AAB16268
 CC given in the present invention.

SQ Sequence 415 BP; 120 A; 85 C; 98 G; 112 T; 0 other;

Query Match Best Local Similarity 58.8%; Score 19.4; DB 21; Length 415;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 CCACATGATGTCGCAAGCGCTGCGTATATA 30
 DB 156 CCATATGGTGGCAAGCGCTTACGACAGATA 128

RESULT 3

ID AAF15168/c standard; cDNA; 634 BP.

AC AAF15168;

DT 13-MAR-2001 (first entry)

DE Trichoderma reesei EST SEQ ID NO:7691.

KW Multiple gene expression; filamentous fungal cell; EST;

KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;

OS metabolic pathway; engineering; catabolic pathway engineering; ss.

PN Trichoderma reesei.

PD WO200056762-A2.

PF 28-SEP-2000.

PR 22-MAR-2000; 2000WO-US07781.

PR 22-MAR-1999; 99US-0273623.

PA (NOVO) NOVO NORDISK BIOTECH INC.

XX (NOVO) NOVO NORDISK AS.

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IC, Olsen PB;

DR WPI; 2000-594572/56.

PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags

PS Claim 89; Page 3104; 3161bp; English.
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring

CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.

SQ Sequence 634 BP; 144 A; 177 C; 136 G; 144 T; 33 other;

Query Match 58.2%; Score 19.2; DB 21; Length 634;
Best Local Similarity 75.0%; Pred. No. 90;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DY 1 GCCACATGAGTGGCAAGCGCTGTGATACC 32
DB || || || || || || || || || || || ||
83 GCACCATFAGCAGCACGCCGTCCTGTAGAAC 52

RESULT 4
AAFO7727
ID AAF07727 standard; CDNA; 906 BP.
XX AAF07727;
XX 13-MAR-2001 (first entry)
DE Fusarium venenatum EST SEQ ID NO:250.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Fusarium venenatum.
XX
PN WO200056762-A2.
PD 28-SEP-2000.
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
DR WPI: 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 86; Page 478-479; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes

CC		in a first filamentous fungal (FF) cell relative to expression of the
CC		same genes in one or more second filamentous fungal cells. Monitoring
CC		the global expression of genes from FF cells allows the production
CC		potential of the microorganisms to be improved. New genes may be
CC		discovered, possible functions of unknown open reading frames can be
CC		identified and gene copy number variation and stability can be
CC		monitored. The expression of genes can be used to study how FF cells
CC		adapt to changes in culture conditions, environmental stress, spore
CC		morphogenesis, recombination, metabolic or catabolic pathway
CC		engineering. Using ESTs provides several advantages over genomic or
CC		a random cDNA clones including elimination of redundancy as one spot on an
CC		array equals one gene or open reading frame, and organisation of the
CC		microarrays based on function of the gene products to facilitate
CC		analysis of the results. AA07478 to AAF11247 represents ESTs from
CC		Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC		niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC		AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC		all specifically claimed in the present invention.
xx		
SQ	Sequence 906 BP; 203 A; 271 C; 208 G; 200 T; 24 other;	
OY	1 GCCACATGAGTGGCAGCGCTGTGTAACC G 33	
Dd	370 GTCTCANNAAGCCGCAGAGACTGTGCTACCG 402	
	Query Match 58.2%; Score 19.2; DB 21; Length 906;	
	Best Local Similarity 72.7%; Pred. No. 94;	
	Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0,	
RESULT 5		
ID	AA565203/c	
XX	AA565203 standard; cDNM: 1284 BP.	
XX	AA565203;	
XX	13-FEB-2002 (first entry)	
DE	DNA encoding novel human diagnostic protein #1007.	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KX	food supplement; medical imaging; diagnostic; genetic disorder; ss.	
OS	Homo sapiens.	
PN	WO200175067-A2.	
XX	11-OCT-2001.	
PD	30-MAR-2001; 2001WO-US08631.	
PF	31-MAR-2000; 2000US-0540217.	
PR	23-AUG-2000; 2000US-0649167.	
XX	(HYSE-) HKSEQ INC.	
PA		
PI	Dmanac RT, Liu C, Tang YT;	
DR	WPI: 2001-639362/73.	
P	P-PSTDB; ABG01016.	
XX	New isolated polynucleotide and encoded polypeptides, useful in	
PT	diagnostics, forensics, gene mapping, identification of mutations	
PT	responsible for genetic disorders or other traits and to assess	
PT	biodiversity	
XX		
PS	Claim 1; SEQ ID NO 1007; 103pp; English.	
XX	The invention relates to isolated polynucleotide (I) and	
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,	
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome	
CC	and gene mapping, and in recombinant production of (II). The	
CC	polynucleotides are also used in diagnostics as expressed sequence tags	

for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

Sequence 1284 BP; 336 A; 308 C; 329 G; 311 T; 0 other;

Query Match 58.2%; Score 19.2; DB 23; Length 1284;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 2 CCACATGATGCGAAGCGCTGCTGATACCG 33
Db 50 CCACATGCAAGCGAAGCGATGATGATCCG 19

RESULT 6
AA222700 standard; DNA; 5660 BP.

AC AA22700;

DT 15-FEB-2000 (first entry)

DE Pseudomonas fluorescens ABC transporter cassette operon I.

KW ATP-binding cassette; transporter; operon; LipBPD; Serratia marcescens;
KW microbe; protein secretion; ds.

OS Pseudomonas fluorescens.

Key Location/Qualifiers

FT CDS 778..2529

FT CDS 2526..3860

FT CDS 3863..5197

FT CDS /product= "component C"

FT CDS /tag= a

FT CDS /product= "component A"

FT CDS /tag= b

FT CDS /product= "component B"

FT CDS /tag= c

FT CDS /product= "component C"

FT CDS /tag= a

FT CDS /product= "component A"

FT CDS /tag= b

FT CDS /product= "component B"

FT CDS /tag= c

FT CDS /product= "component C"

operon from Pseudomonas fluorescens strain 33. The operon comprises 3 genes where the termination codon of the first gene overlaps with the initiation codon of the second gene. The sequences of the encoded proteins have 60, 44 and 46% homology respectively to the LipBPD proteins from *Serratia marcescens*. The novel gene and protein can give or increase the ability of a microbe for secreting a protein.

Sequence 5660 BP; 1131 A; 1817 C; 1736 G; 976 T; 0 other;

Query Match 58.2%; Score 19.2; DB 20; Length 5660;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 GCCACATGATGCGAAGCGCTGCTGATACCG 32
Db 365 GCCTGACGATGCGAAGCGATGCTGTTTACC 396

RESULT 7
ABN18387/C standard; cDNA; 379 BP.

AC ABN18387;

DT 24-JUN-2002 (first entry)

DE Human ORFX polynucleotide sequence SEQ ID NO:5251.

KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis; gene; ss.

OS Homo sapiens.

PN NO200192523-A2.

PD 06-DEC-2001.

PF 29-MAY-2001; 2001MO-US10836.

PR 30-MAY-2000; 2000US-206132P.

PR 29-AUG-2000; 2000US-228716P.

PA (CURA-) CURAGEN CORP.

PI Shinketsu RA, Leach MD;

PI WPI: 2002-106308/14.

PI P-PSDB; ABP02635.

PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders

PS Disclosure; SEQ ID 5251; 1037pp; English.

XX The present invention describes substantially purified human proteins
XX (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
XX in the specification). ABN15762 to ABN27252 encode the human ORFX
XX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
XX treating or preventing a pathology associated with an ORFX-associated
XX disorder in humans, and in the manufacture of a medicament for treating a
XX syndrome associated with ORFX-associated disorder. ORFX polynucleotide
XX sequences can be used in gene therapy. ORFX sequences can be used in the
XX treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
XX psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
XX osteoarthritis, neurodegenerative disorders, disorders related to organ
XX transplantation, cardiovascular diseases, diabetes mellitus, systemic

This sequence represents an ATP-binding cassette (ABC) transporter

CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORF proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 379 BP; 80 A; 141 C; 94 G; 62 T; 2 other;

Query Match 57.6%; Score 19; DB 24; Length 379;
 Best Local Similarity 81.5%; Pred. No. 1e+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 CCACATGAGTGGCAGGCGCTGCTGCA 28
 ||||| ||||| ||||| |||||
 DB 78 CCACATGCGGTGGTGGGCGGTGCA 52

RESULT 8

AAS83486/C
 ID AAS83486 standard; cDNA; 841 BP.

AC AAS83486;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #19290.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG19299.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

PS Claim 1; SEQ ID NO 19290; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 841 BP; 191 A; 273 C; 224 G; 153 T; 0 other;

Query Match 57.0%; Score 18.8; DB 23; Length 841;
 Best Local Similarity 76.7%; Pred. No. 1.4e+02;
 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 3 CACATGAGTGGCAGGCGCTGCTGATACC 32
 ||||| ||||| ||||| |||||
 DB 210 CAGGGGAGTGGCTGCTGCTGCTGATACC 181

RESULT 9

ABN73219/C
 ID ABN73219 standard; cDNA; 472 BP.

AC ABN73219;

DT 03-JUL-2002 (first entry)

DE Bovine embryonic germ (EG) cell cDNA EST 000128a CONTIG 79.

XX Bovine; Bos taurus; EST; expressed sequence tag; totipotence;

KW development; gene; ss.

OS Bos taurus.

PN WO200194550-A2.

PD 13-DEC-2001.

PF 07-JUN-2001; 2001WO-US18576.

PR 07-JUN-2000; 2000US-209874P.

PR 06-JUN-2001; 2001US-0876143.

PA (INF1-) INFGEN INC.

PI Elertsen KJ, Pfister-genskow M, Childs L;

DR WPI; 2002-351289/38.

PT An expressed sequence tag (EST), the expression of which, or its
 PT complementary sequence, in a cell identifies the cell as a
 PT developmentally competent or incompetent cell

PS Example 16; Page 129-130; 584pp; English.

CC The present invention describes an expressed sequence tag (EST), where
 CC the EST is an isolated, enriched, or purified nucleic acid sequence
 CC representing all or part of a gene, the expression of which, or its
 CC complementary sequence, in a cell identifies the cell as a
 CC developmentally competent or incompetent cell. Molecules which induce
 CC developmental competence in a cell line are useful for inducing
 CC totipotence in one or more cells. Molecules which induce developmental
 CC incompetence in a cell line are useful for preventing a full term
 CC pregnancy in an animal and inhibiting totipotence. The molecules are
 CC also useful for treating a disease in an animal by inducing development
 CC of one or more cells of the animal into a specific cell type. The
 CC present sequence represents a bovine EST which is given in the
 CC exemplification of the present invention.

Sequence 472 BP; 110 A; 132 C; 169 G; 57 T; 4 other;

Query Match 56.48; Score 18.6; DB 24; Length 472;

Best Local Similarity 72.78; Pred. No. 1.6e+02;

Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 402 GCCCATGCTGTCAGGCGCTGTGTACTG 370

1 GCCACATGATGTCAGGCGCTGTGTACTG 33

RESULT 10

AAC3442/C

ID AAC3442 standard; DNA; 973 BP.

XX AAC3442;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 3049.

KM Hybridisation assay; genetic mapping; gene expression control;

KM protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 29-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0128845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130444.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134378.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 21-JUN-1999; 99US-0139750.

PR 22-JUN-1999; 99US-0139817.

PR 23-JUN-1999; 99US-0139859.

PR 24-JUN-1999; 99US-0140353.

PR 28-JUN-1999; 99US-0140354.

PR 29-JUN-1999; 99US-0140695.

PR 30-JUN-1999; 99US-0140823.

PR 01-JUL-1999; 99US-0140991.

PR 01-JUL-1999; 99US-0141287.

PR 02-JUL-1999; 99US-0141842.

PR 06-JUL-1999; 99US-0142055.

PR 08-JUL-1999; 99US-0142390.

PR 09-JUL-1999; 99US-0142803.

PR 12-JUL-1999; 99US-0142920.

PR 13-JUL-1999; 99US-0142977.

PR 14-JUL-1999; 99US-0143542.

PR 15-JUL-1999; 99US-0143624.

PR 16-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 19-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 20-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 21-JUL-1999; 99US-0144632.

PR 21-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144884.

PR 22-JUL-1999; 99US-0145086.

PR 22-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 23-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 26-JUL-1999; 99US-0145224.

PR 27-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 28-JUL-1999; 99US-0145919.

PR 02-AUG-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 03-AUG-1999; 99US-0146389.

PR 04-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 05-AUG-1999; 99US-0147302.

PR 06-AUG-1999; 99US-0147192.

PR 06-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 09-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151338.
PR	01-SEP-1999;	99US-0151330.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	29-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	56.4%	Score 18.6	DB 21	Length 973
Best Local Similarity	72.7%	Pred. No. 1.7e+02		
Matches 24	Conservative 0	Mismatches 9	Indels 0	Gaps 0
0y	1	gccacatgacgtgcgaagcgctctgtgtataccg	33	.

Db 268 GCACGAGCAGAGCGACAGCTGTGTATCCAG 236
 RESULT 11
 AAS83436/c
 ID AAS83436 standard; cDNA: 1122 BP.
 AC AAS83436;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #19240.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 RW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PA 23-AUG-2000; 2000US-0649167.
 PI (HYSE-) HYSEQ INC.
 PL Drmanac RT, Liu C, Tang YT;
 DR WPI: 2001-639362/73.
 P-PSDB; ABG19249.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS Claim 1; SEQ ID NO 19240; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_prtl_sequences.
 XX Sequence 1122 BP; 216 A; 400 C; 309 G; 196 T; 1 other;

Query Match	56.4%	Score 18.6	DB 23	Length 1122
Best Local Similarity	72.7%	Pred. No. 1.8e+02		
Matches 24	Conservative 0	Mismatches 9	Indels 0	Gaps 0
OY	1	GCACATGAGTGCAGGCGCTGCTGATACCG	33	
Db	1074	GCCACAGAGAGCGGCTGGGCTCTTTATACCG	1042	

RESULT 12

ABN85733/c
ID ABN85733 standard; DNA: 249487 BP.

AC ABN85733;
XX

DT 30-SEP-2002 (first entry)
XX

DE Mouse genomic region containing the Itprcs gene SEQ ID NO 3.
XX

KW Mouse; Itprcs; taste; cell signalling; TC-ICS; food; pharmaceutical;
XX

KW taste cell-specific ion channel subunit; gene; ds.
XX

OS Mus sp.
XX

PN WO200254069-A1.
XX

PD 11-JUL-2002.
XX

PF 26-DEC-2001; 2001WO-US49808.
XX

PR 29-DEC-2000; 2000US-259379P.
XX

PR 21-DEC-2001; 2001US-0026188.
XX

PA (REGC.) UNIV CALIFORNIA.
XX

PI Zuker CS, Zhang Y;
XX

DR WPI; 2002-583632/62.
XX

XX
XX

PT Identifying modulators of taste signaling in taste cells for use in
XX

PT food and pharmaceutical industries to customize and regulate taste, by
XX

PT determining effect of the compound on a taste cell-specific ion channel
XX

PT subunit
XX

XX
XX

PS Disclosure; Page 64-207; 306pp; English.
XX

XX
XX

CC The invention relates to identifying (M1) a compound that modulates taste
XX

CC signaling in taste cells, by contacting the compound with a eukaryotic
XX

CC host cell or cell membrane which expresses a taste cell-specific ion
XX

CC channel subunit (TC-ICS), and determining a functional effect of the
XX

CC compound upon a transmembrane ion flux of a predetermined ion,
XX

CC identifying a compound that modulates taste signaling in taste cells.
XX

CC (M1) is useful for identifying a compound that binds to a taste cell
XX

CC specific ion channel subunit and for modulating taste signaling in taste
XX

CC cells of a mammal, in particular a human. Modulators identified by (M1)
XX

CC e.g. as additives to food or pharmaceutical industries to customize taste,
XX

CC are used by the food and pharmaceutical industries to customize taste,
XX

CC e.g. as additives to food or medicine so that the food or medicine tastes
XX

CC different to the subject who ingests it. Bitter medicines can be made to
XX

CC taste less bitter and sweet substance can be enhanced. The modulators are
XX

CC useful for pharmacological and genetic modulation of taste signaling
XX

CC pathways. The taste modulators can be directly administered to mammalian
XX

CC subjects for modulation of taste in vivo. The present sequence is that of
XX

CC the mouse genomic region containing the Itprcs gene of the invention.
XX

XX
XX

Sequence 249487 BP: 64808 A; 61913 C; 61498 G; 61236 T; 32 other;
XX

Query Match
XX

Best Local Similarity 56.4%; Score 18.6; DB 24; Length 249487;
XX

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX

DB 25090 ACATGAGTGGCAGCGCTCTGCGA 28
XX

ACATGAGTGGCAGCGCTCTGCGA 28
XX

Result 13
XX

AAA96719/c
XX

AAA96719 standard; DNA: 1131 BP.
XX

AC AAA96719;

DT 19-FEB-2001 (first entry)
XX

DE Polynucleotide isolated from lymph node stromal cells of fsn -/- mice.
XX

KW Lymph node stromal cell; fsn -/- mice; inflammatory disorder;
XX

KW immune system disorder; cancer; viral infection; HIV infection;
XX

KW blood vessel growth; tumour necrosis factor disorder; arthritis;
XX

KW inflammatory bowel disease; fibroblast growth factor-mediated disorder;
XX

OS cardiac failure; ss.
XX

XX
XX

Key Location/Qualifiers
XX

FT CDS 1..360
XX

FT CDS /*tag= a
XX

PN WO200058463-A1.
XX

PD 05-OCT-2000.
XX

PF 18-FEB-2000; 2000WO-NZ00015.
XX

PR 25-MAR-1999; 99US-0276268.
XX

PR 26-AUG-1999; 99US-0383586.
XX

PA (GENE-) GENESIS RES & DEV CORP LTD.
XX

PI Strachan L, Sleeman M, Abernethy N, Orrust R, Kumble KD;
XX

PI Murison JG;
XX

DR WPI; 2000-664924/64.
XX

DR P-PSDB; AAB19109.
XX

XX
XX

PT Polypeptide expressed in mammalian fsn -/- lymph node stromal cells,
XX

PT useful for modulating growth of blood cells, for treating inflammatory
XX

PT and tumour necrosis factor-mediated disorders, cancer and viral
XX

PT disorders
XX

PS Claim 4; Page 35; 75pp; English.
XX

XX
XX

CC The present sequence represents a polynucleotide sequence which is
XX

CC isolated from lymph node stromal cells of fsn -/- mice. The
XX

CC polynucleotides and their polypeptides are useful for treating an
XX

CC inflammatory disorder, disorder of immune system and cancer selected
XX

CC from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
XX

CC viral disorder, in particular HIV infection and for modulating the
XX

CC growth of blood vessels. The polypeptides are useful for treating a
XX

CC tumour necrosis factor (TNF) mediated disorder, such as those selected
XX

CC from arthritis, inflammatory bowel disease and cardiac failure and a
XX

CC to determine biological activity, to raise antibodies, to isolate
XX

CC corresponding ligands or receptors, to quantify levels of protein or
XX

CC and in compositions for the treatment of skin, connective tissue and
XX

CC immune system diseases. The polynucleotide is useful as marker for
XX

CC tissue, as a chromosome marker or tags in the identification of a
XX

CC genetic disorder.
XX

Sequence 1131 BP: 208 A; 343 C; 342 G; 238 T; 0 other;
XX

Query Match
XX

Best Local Similarity 55.8%; Score 18.4; DB 21; Length 1131;
XX

Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX

DB 435 GCCAGATGAGTGGCAGCGAGTCTGCGA 408
XX

GCCAGATGAGTGGCAGCGAGTCTGCGA 408
XX

Result 14
XX

AAF93747
XX

ID AAF93747 standard; cDNA; 3425 BP.
 AC AAF93747;
 DT 23-MAY-2001 (first entry)
 DE Human cDNA encoding a membrane or secretory protein clone PSEC0007.
 XX
 KM Human; secretory protein; membrane protein; vaccine; gene therapy;
 KM rheumatoid arthritis; diabetes; ss.
 OS Homo sapiens.
 PN EP1067182-A2.
 PD 10-JAN-2001.
 PF 07-JUL-2000; 2000EP-0114090.
 PR 08-JUL-1999; 99JP-0194179.
 PR 11-JAN-2000; 2000JP-0118775.
 PR 02-MAY-2000; 2000JP-0183766.
 PA (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 DR WPI: 2001-093989/11.
 DR P-PSDB: AAB88320.
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development -
 PS Claim 1: SEQ ID 7; 609pp + CD ROM; English.
 XX
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by
 CC AAB88317 - AAB88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies (agonists and antagonists) of expression and
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
 CC (ELISA). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.
 CC
 SQ Sequence 3425 BP; 800 A; 840 C; 917 G; 868 T; 0 other;
 XX
 Query Match 55.8%; Score 18.4; DB 22; Length 3425;
 Best Local Similarity 78.6%; Pred. No. 2.5e+02;
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 OY 1 GCCACATGAGTGGCAAGGCTGCTGCTGA 28
 DB 1891 GCCACCTGTGTGGCAAGGCTGCTGCTGA 1918

RESULT 15
 AA248806

ID AA248806 standard; cDNA; 518 BP.
 AC AA248806;
 DT 21-MAR-2000 (first entry)
 DE Rice inositol 1,3,4-triphosphate 5/6-kinase coding sequence.
 XX
 KM Inositol 1,3,4-triphosphate 5/6-kinase; phytic acid; genetic mapping;
 KM myo-inositol 1,2,3,4,5,6-hexaphosphate; biosynthetic enzyme; phytate;
 KM animal feed; ss.
 OS Oryza sativa.
 PN WO9955879-A1.
 PD 04-NOV-1999.
 PF 22-APR-1999; 99WO-US08790.
 PR 24-APR-1998; 98US-0082960.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 PI Cahoon RE, Carlson TJ, Hiltz WD, Pearlstein RM;
 DR WPI: 2000-072179/06.
 DR P-PSDB: AAV59422.
 PT Novel phytic acid biosynthetic enzymes used to alter the level of the
 PT enzyme in transgenic plants -
 PS Claim 2: Page 45; 63pp; English.
 XX
 CC This sequence encodes an inositol 1,3,4-triphosphate 5/6-kinase (ITK)
 CC protein of the invention. ITK is a phytic acid (myo-inositol
 CC 1,2,3,4,5,6-hexaphosphate) biosynthetic enzyme. The ITK enzymes of the
 CC invention may be prepared recombinantly and used to raise antibodies,
 CC which are used for detecting the enzymes in situ in cells or in vitro
 CC in cell extracts. The polynucleotides of the invention may be used to
 CC create transgenic plants in which the ITK levels are present at higher
 CC or lower levels than normal, or in cell types or developmental processes
 CC where they are not normally found. This would alter the level of
 CC 1,3,4-triphosphate 5/6 kinase found in those cells. In addition, it may
 CC be desirable to eliminate expression of ITK genes for certain
 CC applications. The polynucleotides also serve as a source of probes and
 CC primers, which are useful for genetic mapping, as markers for traits
 CC linked to those genes, and to isolate homologous sequences from other
 CC species, as well as for physical mapping. Decreased amounts of phytic
 CC acid and increased amounts of available phosphate in animal feeds such as corn
 CC would lead to improved feed efficiency. The proteins of the present
 CC invention lead to a better understanding of the phytic acid biosynthesis
 CC pathway, allowing it to be exploited for commercial uses, e.g. in animal
 CC feeds.
 CC
 SQ Sequence 518 BP; 120 A; 114 C; 117 G; 159 T; 8 other;
 XX
 Query Match 55.2%; Score 18.2; DB 21; Length 518;
 Best Local Similarity 87.0%; Pred. No. 2.4e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 5 CATGAGTGGCAAGGCTGCTGCTG 27
 DB 198 CATGAGTGGCAAGGCTGCTGCTG 220

Search completed: July 10, 2003, 19:52:39
 Job time : 72.6337 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 14.7573 Seconds
(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631B-5

Perfect score: 33

Sequence: 1 gccacatgagtggaagcgctcgtgataccg 33

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: Issued_Patents_NA.*
2: /cgn2_6/ptodata/1/1na/5A_COMB.seq.*
3: /cgn2_6/ptodata/1/1na/5B_COMB.seq.*
4: /cgn2_6/ptodata/1/1na/6A_COMB.seq.*
5: /cgn2_6/ptodata/1/1na/6B_COMB.seq.*
6: /cgn2_6/ptodata/1/1na/Backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.4	55.8	320	4	US-08-651-155B-214
2	18.4	55.8	1131	4	US-09-383-586-7
3	18.2	55.2	1740	2	US-08-471-025-37
4	18.2	55.2	7785	2	US-08-276-967-1
5	17.8	53.9	617	4	US-09-385-982-396
6	17.8	53.9	1279	2	US-08-872-719-1
7	17.8	53.9	1279	3	US-08-957-302A-11
8	17.8	53.9	1279	4	US-09-336-890-1
9	17.8	53.9	1279	4	US-09-542-403-11
10	17.8	53.9	1279	4	US-09-668-499-1
11	17.8	53.9	1395	3	US-08-957-302A-3
12	17.8	53.9	1395	4	US-09-542-403-3
13	17.8	53.9	1575	3	US-08-957-302A-1
14	17.8	53.9	1575	4	US-09-542-403-1
15	17.6	53.3	949	4	US-09-221-017B-615
16	17.6	53.3	2532	1	US-07-671-376C-4
17	17.6	53.3	19227	3	US-09-090-793-13
18	17.6	53.3	40138	3	US-09-090-793-12
19	17.4	52.7	436	1	US-08-686-878A-19
20	17.4	52.7	456	1	US-09-175-928-19
21	17.4	52.7	2805	4	US-08-959-004-6
22	17.2	52.1	4403765	4	US-09-103-840A-2
23	17.2	52.1	4411529	4	US-09-103-840A-1
24	17	51.5	51	1	US-08-396-218-3
25	17	51.5	51	1	US-08-760-116-3
26	17	51.5	59	1	US-08-396-218-4
27	17	51.5	59	1	US-08-760-116-4

c 28	17	51.5	318	1	US-08-259-372A-15	Sequence 15, Appl
c 29	17	51.5	318	1	US-08-468-671-15	Sequence 15, Appl
c 30	17	51.5	594	4	US-09-385-982-397	Sequence 397, App
c 31	17	51.5	620	1	US-08-328-055-2	Sequence 2, Appl
c 32	17	51.5	642	1	US-08-480-753-5	Sequence 5, Appl
c 33	17	51.5	642	3	US-09-041-889-10	Sequence 10, Appl
c 34	17	51.5	642	3	US-08-837-058-10	Sequence 10, Appl
c 35	17	51.5	785	1	US-08-328-055-5	Sequence 5, Appl
36	17	51.5	1446	1	US-08-596-024-5	Sequence 5, Appl
37	17	51.5	1446	4	US-09-020-818-5	Sequence 5, Appl
38	17	51.5	1446	4	US-08-907-740-5	Sequence 5, Appl
39	17	51.5	2081	2	US-09-096-982-7	Sequence 7, Appl
40	17	51.5	2081	2	US-08-653-650A-7	Sequence 7, Appl
41	17	51.5	2910	4	US-09-206-942-31	Sequence 31, Appl
42	17	51.5	2928	4	US-09-206-942-29	Sequence 29, Appl
43	17	51.5	3196	2	US-09-096-982-4	Sequence 4, Appl
44	17	51.5	3196	2	US-08-653-650A-4	Sequence 4, Appl
45	17	51.5	13875	2	US-08-734-344-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-651-155B-214
Sequence 214, Application US/08651155B
Patent No. 6365401
GENERAL INFORMATION:
APPLICANT: Mahan Dr., Michael J.
APPLICANT: Conner Mr., Christopher P.
APPLICANT: Hiethoff Mr., Douglas M.
TITLE OF INVENTION: METHOD AND PROBS FOR THE IDENTIFICATION
TITLE OF INVENTION: OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST
NUMBER OF SEQUENCES: 255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chrisman, Bynum & Johnson, P.C.
STREET: 1900 Fifteenth Street
CITY: Boulder
STATE: CO
COUNTRY: USA
ZIP: 80302
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651,155B
FILING DATE: 17-MAY-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Petersen Mr., Steven C.
REGISTRATION NUMBER: 36,238
REFERENCE/DOCKET NUMBER: 17060.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/546-1300
TELEFAX: 303/449-5426
TELEX: ABA1475
INFORMATION FOR SEQ ID NO: 214:
SEQUENCE CHARACTERISTICS:
LENGTH: 320 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: YES
US-08-651-155B-214
Query Match 55.8%; Score 18.4; DB 4; Length 320;
Best Local Similarity 78.6%; Pred. No. 19;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY

6 ATGAGTGCAGAGCGCTGCTGATACCG 33

DB

196 ATGGCTGGCATGAGCTGCGCAACCG 223

RESULT 2

US-09-383-586-7/c

Sequence 7, Application US/09383586

Patent No. 6242419

GENERAL INFORMATION:

APPLICANT: Strachan, Lotina

APPLICANT: Sleeman, Matthew

APPLICANT: Abernethy, Nevin

APPLICANT: Onrust, Rene

APPLICANT: Kumdie, Anand

APPLICANT: Mutison, Greg

TITLE OF INVENTION: Compounds isolated from stromal cells

FILE REFERENCE: 11000.1037c1

CURRENT APPLICATION NUMBER: US/09/383,586

NUMBER OF SEQ ID NOS: 38

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 7

LENGTH: 1131

TYPE: DNA

ORGANISM: Mouse

US-09-383-586-7

Query Match

Best Local Similarity 78.68; Score 18.4; DB 4; Length 1131;

Matches 32; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY

1 GCCACATGAGTGCAGAGCGCTGCTGTA 28

DB

435 GCCAGATGAGTGCAGAGCGCTGCTGCGCA 408

RESULT 3

US-08-471-025-37/c

Sequence 37, Application US/08471025

Patent No. 598561

GENERAL INFORMATION:

APPLICANT: Paolletti, Enzo

APPLICANT: Fischer, Laurent

TITLE OF INVENTION: RECOMBINANT POXVIRUS - CALICIVIRUS

TITLE OF INVENTION: [RABBIT HEMORRHAGIC DISEASE VIRUS (RHDV)] COMPOSITIONS AND

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESS: Curtis, Morris & Safford, P.C.

STREET: 530 Fifth Avenue, 25th Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,025

FILING DATE: 06-JUN-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Frommer, William S.

REGISTRATION NUMBER: 25,506

REFERENCE/DOCKET NUMBER: 454310-2650

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333

TELEFAX: (212) 840-0712

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:

LENGTH: 1740 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-471-025-37

Query Match

Best Local Similarity 55.28; Score 18.2; DB 2; Length 1740;

Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 CCACATGAGTGCAGAGCGCTGCTGATACC 32

DB 1020 GCACATGACAGGAGCGCTGCTGATACC 990

RESULT 4

US-08-276-967-1

Sequence 1, Application US/08276967

Patent No. 5851817

GENERAL INFORMATION:

APPLICANT: Hardy, Daniel M.

APPLICANT: Garbers, David L.

TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESS: Arnold, White & Durkee

STREET: P. O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/276,967

FILING DATE: Submitted Herewith

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Kitchell, Barbara S.

REGISTRATION NUMBER: 33,928

REFERENCE/DOCKET NUMBER: UTSD:418/KIT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 713-787-1400

TELEFAX: 713-789-2679

TELEX: 79-0924

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 7785 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-276-967-1

Query Match

Best Local Similarity 55.28; Score 18.2; DB 2; Length 7785;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 10 GTGGCAGAGCGCTGCTGATACC 32

DB 509 GTGGCAGAGCGCTGCTGATACC 531

RESULT 5

US-09-385-982-396

Sequence 396, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCDA-260XX
CURRENT FILING DATE: 1999-08-30
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 396
LENGTH: 617
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(617)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-396

Query Match
Best Local Similarity 53.9%; Score 17.8; DB 4; Length 617;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCCACATGAGTGGCAGCGCTGTGTGAT 29
DB 12 GCCACTGAGTGGCAGCGCTGTGTGAT 40

RESULT 6

US-08-872-719-1

Sequence 1, Application US/08872719

Patent No. 5919643

GENERAL INFORMATION:

APPLICANT: Kelley, Mark R.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE USE OF APURINIC/APYRIMIDINIC

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARNOLD, WHITE & DURKEE

STREET: P.O. BOX 4433

CITY: HOUSTON

STATE: TEXAS

COUNTRY: USA

ZIP: 77057-2196

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/872,719

FILING DATE: CONCURRENTLY HERewith

CLASSIFICATION: 436

ATTORNEY/AGENT INFORMATION:

NAME: Highlander, Steven L.

REGISTRATION NUMBER: 37,642

REFERENCE/DOCKET NUMBER: INDY:012P21

TELEPHONE: (512) 418-3000

TELEFAX: (713) 789-2679

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1279 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-872-719-1

Query Match
Best Local Similarity 53.9%; Score 17.8; DB 2; Length 1279;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCCACATGAGTGGCAGCGCTGTGTGAT 29
DB 821 GCCACAGAGCGCCAGCGCTGTGTGAT 849

RESULT 7

US-08-957-302A-11

Sequence 11, Application US/08957302A

Patent No. 6046036

GENERAL INFORMATION:

APPLICANT: Williams, David

TITLE OF INVENTION: DNA Sequences Encoding Fusions of DNA

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARNOLD, WHITE & DURKEE

STREET: P.O. BOX 4433

CITY: HOUSTON

STATE: TX

COUNTRY: US

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/957,302A

FILING DATE: Concurrently Herewith

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Highlander, Steven L.

REGISTRATION NUMBER: 37,642

REFERENCE/DOCKET NUMBER: INDY:005

TELEPHONE: (512) 418-3000

TELEFAX: (713) 789-2679

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 1279 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-957-302A-11

Query Match
Best Local Similarity 53.9%; Score 17.8; DB 3; Length 1279;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCCACATGAGTGGCAGCGCTGTGTGAT 29
DB 821 GCCACAGAGCGCCAGCGCTGTGTGAT 849

RESULT 8

US-09-336-890-1

Sequence 1, Application US/09336890

Patent No. 6190661

GENERAL INFORMATION:

APPLICANT: Kelley, Mark R.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE USE OF

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARNOLD, WHITE & DURKEE

STREET: P.O. BOX 4433

CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77057-2198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/336,890
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/872,719
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: INDY:012P21
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1279 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-336-890-1

Query Match 53.9%; Score 17.8; DB 4; Length 1279;
Best Local Similarity 75.9%; Pred. No. 45;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCCACATGATGCGCAAGCGCTGTGAT 29
||||| ||| ||||| ||||| ||
DB 821 GCCACAAGAGCGCCAGGCTTCGGGGAAT 849

RESULT 9
US-09-542-403-11
Sequence 11, Application US/09542403
Patent No. 6252048
GENERAL INFORMATION:
APPLICANT: Kelley, Mark
TITLE OF INVENTION: DNA Sequences Encoding Fusions of DNA
TITLE OF INVENTION: Repair Proteins and Uses Thereof
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSER: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TX
COUNTRY: US
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/542,403
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/957,302
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: INDY:005

TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1279 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-542-403-11

Query Match 53.9%; Score 17.8; DB 4; Length 1279;
Best Local Similarity 75.9%; Pred. No. 45;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCCACATGATGCGCAAGCGCTGTGAT 29
||||| ||| ||||| ||||| ||
DB 821 GCCACAAGAGCGCCAGGCTTCGGGGAAT 849

RESULT 10
US-09-668-499-1
Sequence 1, Application US/09668499
Patent No. 6406917
GENERAL INFORMATION:
APPLICANT: Kelley, Mark R.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE USE OF APURINIC/APYRIMIDIN
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSER: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77057-2198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/668,499
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/872,719
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: INDY:012P21
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1279 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-668-499-1

Query Match 53.9%; Score 17.8; DB 4; Length 1279;
Best Local Similarity 75.9%; Pred. No. 45;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCCACATGATGCGCAAGCGCTGTGAT 29
||||| ||| ||||| ||||| ||
DB 821 GCCACAAGAGCGCCAGGCTTCGGGGAAT 849

RESULT 11
US-08-957-302A-3

Sequence 3, Application US/08957302A
Patent No. 6046036
GENERAL INFORMATION:
APPLICANT: Kelley, Mark
TITLE OF INVENTION: DNA Sequences Encoding Fusions of DNA
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TX
COUNTRY: US
ZIP: 77210-4433
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957.302A
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: INDY:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1395 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-957-302A-3

Query Match 53.9%; Score 17.8; DB 3; Length 1395;
Best Local Similarity 75.9%; Pred. No. 46;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GCCACATGAGTGGCAAGCGCTGCTGTAT 29
||||| ||| | ||||| ||| ||
Db 1137 GCCACAAGAGCGCCAGCGCTTCGGGGAT 1165

RESULT 12
US-09-542-403-3
Sequence 3, Application US/09542403
Patent No. 6252048
GENERAL INFORMATION:
APPLICANT: Kelley, Mark
TITLE OF INVENTION: DNA Sequences Encoding Fusions of DNA
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TX
COUNTRY: US
ZIP: 77210-4433
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/542.403
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/957.302
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: INDY:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1395 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-542-403-3

Query Match 53.9%; Score 17.8; DB 4; Length 1395;
Best Local Similarity 75.9%; Pred. No. 46;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GCCACATGAGTGGCAAGCGCTGCTGTAT 29
||||| ||| | ||||| ||| ||
Db 1137 GCCACAAGAGCGCCAGCGCTTCGGGGAT 1165

RESULT 13
US-08-957-302A-1
Sequence 1, Application US/08957302A
Patent No. 6046036
GENERAL INFORMATION:
APPLICANT: Kelley, Mark
TITLE OF INVENTION: DNA Sequences Encoding Fusions of DNA
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TX
COUNTRY: US
ZIP: 77210-4433
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957.302A
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: INDY:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1575 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-957-302A-1

Query Match 53.9%; Score 17.8; DB 3; Length 1575;
Best Local Similarity 75.9%; Pred. No. 47;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 GCCACATGAGTGGCAGGCGCTGTGTGAT 29
DB 1317 GCCACAGAGCGCCCAAGGCTTCGGGGAAT 1345

RESULT 14

US-09-542-403-1

Sequence 1, Application US/09542403

Patent No. 6252048

GENERAL INFORMATION:

APPLICANT: Kelley, Mark

APPLICANT: Williams, David

TITLE OF INVENTION: DNA Sequences Encoding Fusions of DNA

TITLE OF INVENTION: Repair Proteins and Uses Thereof

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESSES:

ADDRESSEE: ARNOLD, WHITE & DUNKEE

STREET: P.O. BOX 4433

CITY: HOUSTON

STATE: TX

COUNTRY: US

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/542,403

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/957,302

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Highlander, Steven L.

REGISTRATION NUMBER: 37,642

REFERENCE/DOCKET NUMBER: INDY:005

TELEPHONE: (512) 418-3000

TELEFAX: (713) 789-2679

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1575 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-542-403-1

Query Match 53.9%; Score 17.8; DB 4; Length 1575;

Best Local Similarity 75.9%; Pred. No. 47;

Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 GCCACATGAGTGGCAGGCGCTGTGTGAT 29

DB 1317 GCCACAGAGCGCCCAAGGCTTCGGGGAAT 1345

RESULT 15

US-09-221-017B-615

Sequence 615, Application US/09221017B

Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.

TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FASTSEQ for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: MONROY, Gladys H

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELEPHONE: 650-813-5600

TELEFAX: 650-494-0792

INFORMATION FOR SEQ ID NO: 615:

SEQUENCE CHARACTERISTICS:

LENGTH: 949 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: UNKNOWN

ORIGINAL SOURCE:

ORGANISM: PORPHYROMONAS GINGIVALIS

FEATURE:

NAME/KEY: misc.feature

LOCATION: 1...949

US-09-221-017B-615

Query Match 53.3%; Score 17.6; DB 4; Length 949;

Best Local Similarity 71.9%; Pred. No. 53;

Matches 23; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 1 GCCACATGAGTGGCAGGCGCTGTGTGATAC 32

DB 829 GCATATGATTTGCCAAGAGTTTCGGGATAC 860

Search completed: July 10, 2003, 20:27:34

Job time: 34.7573 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:53:16 ; Search time 90.3236 Seconds
(Without alignments)
576.891 Million cell updates/sec

Title: US-10-087-631B-5

Perfect score: 33
Sequence: 1 gccacatgagtgcaagcgctgtgatccg 33

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
- 2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 6: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	33	100.0	33	US-10-087-631B-5	Sequence 5, Appl1
2	22	66.7	241	US-10-087-631B-12	Sequence 12, Appl
3	22	66.7	242	US-10-087-631B-15	Sequence 15, Appl
4	22	66.7	375	US-10-087-631B-11	Sequence 14, Appl
5	22	66.7	943	US-10-087-631B-11	Sequence 11, Appl
6	20	60.6	744	US-09-995-598-85	Sequence 85, Appl
7	20	60.6	744	US-09-995-598-87	Sequence 87, Appl
8	20	60.6	744	US-09-995-598-89	Sequence 89, Appl
9	19.8	60.0	771	US-10-013-173-38	Sequence 38, Appl
10	19.8	60.0	771	US-10-150-762-38	Sequence 38, Appl
11	19.4	58.8	499	US-09-918-995-25110	Sequence 13265, A
12	19.2	58.2	405	US-09-960-352-13265	Sequence 13265, A
13	19	57.6	161280	US-10-144-649A-746	Sequence 746, App
14	18.6	56.4	1197	US-10-156-761-409	Sequence 409, App
15	18.6	56.4	100000	US-10-156-761-15103	Sequence 15103, A
16	18.6	56.4	249487	US-10-026-188-3	Sequence 3, Appl1
17	18.6	56.4	9025608	US-10-156-761-1	Sequence 1, Appl1
18	18.4	55.8	1131	US-09-823-038A-7	Sequence 7, Appl1
19	18.2	55.8	189	US-09-783-590-9470	Sequence 9470, Ap

c 20	18.2	55.2	409	10	US-09-878-574-3877	Sequence 3877, Ap
c 21	18.2	55.2	411	10	US-09-983-965-1205	Sequence 1205, Ap
c 22	18.2	55.2	418	10	US-09-983-965-1359	Sequence 1359, Ap
c 23	18.2	55.2	421	10	US-09-983-965-1445	Sequence 1445, Ap
c 24	18.2	55.2	853	9	US-09-764-891-9060	Sequence 9060, Ap
c 25	18.2	55.2	2265	9	US-10-128-714-2448	Sequence 2448, Ap
c 26	18.2	55.2	2763	9	US-10-128-714-7448	Sequence 7448, Ap
c 27	18.2	55.2	2817	9	US-10-128-714-1448	Sequence 1448, Ap
c 28	18.2	55.2	2817	9	US-10-128-714-6448	Sequence 6448, Ap
c 29	18.2	55.2	4817	9	US-10-128-714-448	Sequence 448, App
c 30	18.2	55.2	4817	9	US-10-128-714-5448	Sequence 5448, App
c 31	18	54.5	593	10	US-09-864-761-12260	Sequence 12260, A
c 32	18	54.5	465237	10	US-09-933-267A-1	Sequence 1, Appl1
c 33	17.8	53.9	617	9	US-09-871-161-396	Sequence 396, Appl
c 34	17.8	53.9	1441	9	US-10-171-581-69	Sequence 69, Appl
c 35	17.8	53.9	1484	10	US-09-925-300-248	Sequence 248, App
c 36	17.8	53.9	2052	9	US-10-081-872-187	Sequence 187, App
c 37	17.8	53.9	3730	10	US-09-880-107-1603	Sequence 1603, Ap
c 38	17.6	53.3	194	10	US-09-974-300-3703	Sequence 3703, Ap
c 39	17.6	53.3	348	10	US-09-878-574-2693	Sequence 2693, Ap
c 40	17.6	53.3	486	9	US-09-918-995-20104	Sequence 20104, A
c 41	17.6	53.3	526	10	US-09-560-863-194	Sequence 194, App
c 42	17.6	53.3	762	9	US-10-151-882-12	Sequence 12, Appl
c 43	17.6	53.3	768	9	US-10-156-761-4422	Sequence 4422, Ap
c 44	17.6	53.3	1312	9	US-09-764-891-5455	Sequence 5455, Ap
c 45	17.6	53.3	1837	10	US-09-925-301-365	Sequence 365, App

ALIGNMENTS

RESULT 1
US-10-087-631B-5
Sequence 5, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087, 631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 33
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: ST650pc probe sequence
US-10-087-631B-5

Query Match 100.0%; Score 33; DB 9; Length 33;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1 gccacatgagtgcaagcgctgtgatccg 33
Db 1 gccacatgagtgcaagcgctgtgatccg 33
RESULT 2
US-10-087-631B-12/C
Sequence 12, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087, 631B
CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin version 3.1
SEQ ID NO 12
LENGTH: 241
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Amplicon derived from QS (pc
US-10-087-631B-12

Query Match 66.7%; Score 22; DB 9; Length 241;
Best Local Similarity 97.1%; Pred. No. 1.3;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GCCACATGAGTGG-CAAGCGCTGTGTGATACCG 33
Db 103 GCCACATGAGTGGCCAAAGCGCTGTGTGATACCG 70

RESULT 3
US-10-087-631B-15/c
Sequence 15, Application US/10087631B
Publication No. US20030054372A1

GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin version 3.1
SEQ ID NO 15
LENGTH: 242
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of artificial sequence: Amplicon derived from ICSJ62
US-10-087-631B-15

Query Match 66.7%; Score 22; DB 9; Length 242;
Best Local Similarity 97.1%; Pred. No. 1.3;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GCCACATGAGTGG-CAAGCGCTGTGTGATACCG 33
Db 103 GCCACATGAGTGGCCAAAGCGCTGTGTGATACCG 70

RESULT 4
US-10-087-631B-14/c
Sequence 14, Application US/10087631B
Publication No. US20030054372A1

GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin version 3.1
SEQ ID NO 14
LENGTH: 375
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: ICSJ620 HCV (HCV specific
OTHER INFORMATION: amplification control having a binding site for S1280 and S1778
US-10-087-631B-14

Query Match 66.7%; Score 22; DB 9; Length 375;
Best Local Similarity 97.1%; Pred. No. 1.3;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GCCACATGAGTGG-CAAGCGCTGTGTGATACCG 33
Db 176 GCCACATGAGTGGCCAAAGCGCTGTGTGATACCG 143

RESULT 5
US-10-087-631B-11/c
Sequence 11, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin version 3.1
SEQ ID NO 11
LENGTH: 943
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: QS (pc) HCV being parallel
US-10-087-631B-11

Query Match 66.7%; Score 22; DB 9; Length 943;
Best Local Similarity 97.1%; Pred. No. 1.4;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GCCACATGAGTGG-CAAGCGCTGTGTGATACCG 33
Db 159 GCCACATGAGTGGCCAAAGCGCTGTGTGATACCG 126

RESULT 6
US-09-995-598-85/c
Sequence 85, Application US/09995598
Patent No. US20020142293A1
GENERAL INFORMATION:
APPLICANT: CARINO, RADU
APPLICANT: GILLOT, SOPHIE
TITLE OF INVENTION: ENTEROVIURAL POLYNUCLEOTIDES, METHODS OF DETECTING ENTEROVIURSE
FILE REFERENCE: 215059US-660-660-0
CURRENT APPLICATION NUMBER: US/09/995,598
CURRENT FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/253741
PRIOR FILING DATE: 2000-11-29
NUMBER OF SEQ ID NOS: 96
SOFTWARE: Patentin version 3.1
SEQ ID NO 85
LENGTH: 744
TYPE: DNA
ORGANISM: Echovirus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(744)
OTHER INFORMATION:
US-09-995-598-85

Query Match 60.6%; Score 20; DB 10; Length 744;
Best Local Similarity 82.1%; Pred. No. 11;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 CATGAGTGGCAAGCGCTGTGTGATACCG 32

Db 436 CATGAGTGGCAGGCTGCTGTGACAC 409

RESULT 7
US-09-995-598-87/c
; Sequence 87, Application US/09995598
; Patent No. US20020142293A1
; GENERAL INFORMATION:
; APPLICANT: CRAINIC, RADU
; APPLICANT: CARO, VALERIE
; APPLICANT: GULLLOT, SOPHIE
; TITLE OF INVENTION: ENTEROVIRAL POLYNUCLEOTIDES, METHODS OF DETECTING ENTEROVIRUSES
; FILE REFERENCE: 215059US-660-660-0
; CURRENT APPLICATION NUMBER: US/09/995,598
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253741
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 87
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Echovirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(744)
; OTHER INFORMATION:
US-09-995-598-87

Query Match 60.6%; Score 20; DB 10; Length 744;
Best Local Similarity 82.1%; Pred. No. 11;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 CATGAGTGGCAAGCGCTGCTGTGATACC 32
Db 436 CATGAGTGGCAGGCTGCTGTGACAC 409

RESULT 8
US-09-995-598-89/c
; Sequence 89, Application US/09995598
; Patent No. US20020142293A1
; GENERAL INFORMATION:
; APPLICANT: CRAINIC, RADU
; APPLICANT: CARO, VALERIE
; APPLICANT: GULLLOT, SOPHIE
; TITLE OF INVENTION: ENTEROVIRAL POLYNUCLEOTIDES, METHODS OF DETECTING ENTEROVIRUSES
; FILE REFERENCE: 215059US-660-660-0
; CURRENT APPLICATION NUMBER: US/09/995,598
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253741
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 89
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Echovirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(744)
; OTHER INFORMATION:
US-09-995-598-89

Query Match 60.6%; Score 20; DB 10; Length 744;
Best Local Similarity 82.1%; Pred. No. 11;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 CATGAGTGGCAAGCGCTGCTGTGATACC 32
|||||

Db 436 CATGAGTGGCAGGCTGCTGTGACAC 409

RESULT 9
US-10-013-173-38
; Sequence 38, Application US/10013173
; Publication No. US20030095977A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen C.
; APPLICANT: Graves, Scott Stoll
; APPLICANT: Schultz, Joanne Elaine
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James A.
; APPLICANT: Reno, John M.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; FILE REFERENCE: 690022.547C1
; CURRENT APPLICATION NUMBER: US/10/013,173
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PRIA3 single chain antibody-genomic Streptavidin
; OTHER INFORMATION: fusion construct
US-10-013-173-38

Query Match 60.0%; Score 19.8; DB 9; Length 771;
Best Local Similarity 77.4%; Pred. No. 14;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CCACATGAGTGGCAAGCGCTGCTGTGATACC 32
Db 54 CAAGATCAGCTGCAAGCGCTGCTGTGATACC 84

RESULT 10
US-10-150-762-38
; Sequence 38, Application US/10150762
; Publication No. US20030103948A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen C.
; APPLICANT: Graves, Scott S.
; APPLICANT: Schultz, Joanne E.
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James A.
; APPLICANT: Reno, John M.
; APPLICANT: Dearstyn, Erica A.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; FILE REFERENCE: 690022.547C2
; CURRENT APPLICATION NUMBER: US/10/150,762
; CURRENT FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PRIA3 single chain antibody-genomic Streptavidin
; OTHER INFORMATION: fusion construct
US-10-150-762-38

Query Match 60.0%; Score 19.8; DB 9; Length 771;
Best Local Similarity 77.4%; Pred. No. 14;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CCACATGAGTGGCAAGCGCTGCTGTGATACC 32
|||||

Db 54 CAAGATCAGCTGCAGAGCTTCTGTTATACC 84

RESULT 11

US-09-918-995-25170
; Sequence 25170, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES

FILE REFERENCE: 20411-756

CURRENT APPLICATION NUMBER: US/09/918, 995

PRIOR FILING DATE: 2001-07-30

PRIOR FILING DATE: 1999-01-20

NUMBER OF SEQ ID NOS: 38054

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 25170

LENGTH: 499

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc-feature

LOCATION: (1)...(499)

OTHER INFORMATION: n = A,T,C or G

US-09-918-995-25170

Query Match

Best Local Similarity 58.8%; Score 19.4; DB 9; Length 499;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db

4 ACATGAGTGCAGAGCGCTGTGATACC 32
455 ACATAGTAAACAAGGCTCTGTGATC 483

RESULT 12

US-09-960-352-13265

; Sequence 13265, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Ningbing

APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

FILE REFERENCE: 16511, 006/37-21(10298)C

CURRENT APPLICATION NUMBER: US/09/960,352

PRIOR FILING DATE: 2001-09-24

NUMBER OF SEQ ID NOS: 15112

SEQ ID NO 13265

LENGTH: 405

TYPE: DNA

ORGANISM: Bos taurus

OTHER INFORMATION: Clone ID: 57-LIB188-003-Q1-E1-G10

US-09-960-352-13265

Query Match

Best Local Similarity 58.2%; Score 19.2; DB 10; Length 405;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db

1 GCCACATGAGTGCAGAGCGCTGTGATACC 32
113 GCCATATGGTGCAGATGCTGCGGCTGCC 144

RESULT 13

US-10-144-649A-746

; Sequence 746, Application US/10144649A

; Publication No. US20030118599A1

; GENERAL INFORMATION:

APPLICANT: Lodes, Michael J.

APPLICANT: Wang, Tongtong

APPLICANT: Fan, Liqun

APPLICANT: Algate, Paul A.

APPLICANT: McNeill, Patricia D.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

FILE REFERENCE: 210121, 475C11

CURRENT APPLICATION NUMBER: US/10/144, 649A

PRIOR FILING DATE: 2002-08-21

NUMBER OF SEQ ID NOS: 749

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 746

LENGTH: 161280

TYPE: DNA

ORGANISM: Homo sapiens

US-10-144-649A-746

Query Match

Best Local Similarity 57.6%; Score 19; DB 9; Length 161280;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db

2 CCACATGAGTGCAGAGCGCTGTGATACC 28
39961 CCACATGAGTGCAGAGCGCTGTGATACC 39987

RESULT 14

US-10-156-761-409/C

; Sequence 409, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

PRIOR FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 409

LENGTH: 1197

TYPE: DNA

ORGANISM: Streptomyces avermitilis

FEATURE:

NAME/KEY: CDS

LOCATION: (1) ..(1197)

US-10-156-761-409

Query Match

Best Local Similarity 56.4%; Score 18.6; DB 9; Length 1197;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db

1 GCCACATGAGTGCAGAGCGCTGTGATACC 33
1019 GCCACATGAGTGCAGAGCGCTGTGATACC 987

RESULT 15

US-10-156-761-15103/C

; Sequence 15103, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

: APPLICANT: ISHIKAWA, JUN
 : APPLICANT: HORIKAWA, HIROSHI
 : APPLICANT: SHIBA, TADAYOSHI
 : APPLICANT: SAKAKI, YOSHIYUKI
 : APPLICANT: HATORI, MASAHIRA
 : TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 : FILE REFERENCE: 249-262
 : CURRENT APPLICATION NUMBER: US/10/156,761
 : CURRENT FILING DATE: 2002-05-29
 : PRIOR APPLICATION NUMBER: JP 2001-204089
 : PRIOR FILING DATE: 2001-05-30
 : PRIOR APPLICATION NUMBER: JP 2001-272697
 : PRIOR FILING DATE: 2001-08-02
 : NUMBER OF SEQ ID NOS: 15109
 : SEQ ID NO.15103
 : LENGTH: 100000
 : TYPE: DNA
 : ORGANISM: Streptomyces avermitilis
 US-10-156-761-15103

Query Match 56.4%; Score 18.6; DB 9; Length 100000;
 Best Local Similarity 72.7%; Pred. No. 71;
 Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GCCACATGAGTGGCAGCGCTGTGTATACCG 33
 ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 81448 GCCAGATGGGGGGGGGGCGCTGCGATGTCG 81416

Search completed: July 11, 2003, 15:01:56
 Job time: 100.324 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:13:45 ; Search time 593.333 Seconds
(without alignments)
900.761 Million cell updates/sec

Title: US-10-087-631b-5

Perfect score: 33

Sequence: 1 gccacatgagtggaagcgctcgtgataccg 33

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estopl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rnd:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22.6	68.5	615	10	AV736417 AV736417
2	22.4	67.9	587	10	AM180058 MGN0121f
3	22	66.7	211	9	AV003500 AV003500
4	22	66.7	593	17	AQ989608 RfC00171
5	21.6	65.5	118	10	AM804244 PM3-DM008
6	21.4	64.8	149	17	AZ922863 SLC05F03

c	7	21.2	64.2	777	12	BE676553	BE676553	602084468
c	8	21	63.6	230	10	BE078191	BE078191	CWO-BP061
c	9	21	63.6	609	10	AV744190	AV744190	AV744190
c	10	21	63.6	778	12	BE784735	BE784735	601473642
c	11	20.8	63.0	538	13	B1296816	B1296816	UI-R-DK0
c	12	20.6	62.4	601	10	AV737560	AV737560	AV737560
c	13	20.6	62.4	602	10	AV744188	AV744188	AV744188
c	14	20.6	62.4	613	10	AV737831	AV737831	AV737831
c	15	20.6	62.4	777	9	A1525664	A1525664	PT1.3.04
c	16	20.6	62.4	960	12	B6112389	B6112389	602281958
c	17	20.4	61.8	501	10	BE375363	BE375363	BE375363
c	18	20.4	61.8	589	14	BQ321112	BQ321112	QVO-CT022
c	19	20.2	61.2	337	9	AA074362	AA074362	Zf82904.s
c	20	20.2	61.2	496	10	AV643583	AV643583	AV643583
c	21	20.2	61.2	618	9	A1855688	A1855688	sc23a06.y
c	22	20.2	61.2	699	9	AA089186	AA089186	ml85h01.f
c	23	20.2	61.2	1062	12	BE799927	BE799927	601587073
c	24	20.2	61.2	484	12	BG886095	BG886095	dad5504.f
c	25	20	60.6	489	17	BH333185	BH333185	CH230-202
c	26	20	60.6	561	9	AL794328	AL794328	AL794328
c	27	20	60.6	605	9	AL794563	AL794563	AL794563
c	28	20	60.6	628	9	AL782209	AL782209	AL782209
c	29	20	60.6	719	17	AQ160089	AQ160089	m9xb00036
c	30	20	60.6	735	9	AA706626	AA706626	ab26d08.s
c	31	19.8	60.0	388	12	BE982337	BE982337	UI-M-CG0P
c	32	19.8	60.0	472	10	BB829349	BB829349	BB829349
c	33	19.8	60.0	494	9	AL499781	AL499781	AL499781
c	34	19.8	60.0	495	10	BE634034	BE634034	uv68b09.y
c	35	19.8	60.0	496	9	A1029512	A1029512	UI-R-C0-1
c	36	19.8	60.0	504	17	BH809575	BH809575	KG07290-5
c	37	19.8	60.0	542	10	BE691497	BE691497	uv68b09.x
c	38	19.8	60.0	600	17	AQ094317	AQ094317	Sheared D
c	39	19.8	60.0	615	14	BM942134	BM942134	UI-M-CG0P
c	40	19.8	60.0	637	17	AQ950415	AQ950415	Sheared D
c	41	19.8	60.0	644	10	BB457668	BB457668	BB457668
c	42	19.8	60.0	744	17	AQ650190	AQ650190	Sheared D
c	43	19.6	59.4	572	10	AV742229	AV742229	AV742229
c	44	19.6	59.4	612	10	AV737807	AV737807	AV737807
c	45	19.6	59.4	682	13	BM397994	BM397994	5009-0-4-

ALIGNMENTS

RESULT 1
LOCUS AV736417/c 615 bp mRNA linear EST 17-OCT-2000
DEFINITION AV736417 CB Homo sapiens CDNA clone CBND805 5', mRNA sequence.
ACCESSION AV736417
VERSION AV736417.1 GI:10853998
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 615)
Zhang, Q., Ye, M., Wu, X., Gu, J., Huang, Q., Zhou, J., Shen, Y., Han, Z.,
Chen, S., Mao, M. and Chen, Z.
Homo sapiens CB library cDNA clones
Unpublished (2000)
Contact: Zhu Chen
Shanghai Institute of Hematology, Rui-jin Hospital
197 Rui-jin II Road, Shanghai 200025, P. R. China
Tel: 86-21-64740490
Fax: 86-21-64743206
Email: mshah@sh.sh.cn
This clone is available at Shanghai Hematology Institute in
Shanghai.
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong.
Location/Qualifiers
1..615
/organism="Homo sapiens"

FEATURES

source

Query Match	68.58;	Score 22.6;	DB 10;	Length 615;
Best Local Similarity	86.28;	Pred. No. 78;		
Matches	25;	Conservative	0;	Mismatches

Qy 1 GCCACATGAGTGGCAAGCGCTCTGTGAT 29
||| ||||| ||| ||||| |||||
Db 539 GCCTCATGAGTGGCAAGACGCTTGTGAT 511

	RESULT 2
AM180058	
LOCUS	
DEFINITION	MgA0121f Mga Library Mycosphaerella graminicola cDNA clone MgA0121
ACCESSION	587 bp mRNA linear EST 17-NOV-1999
	5' mRNA sequence.
AM180058	

KEYWORDS	SOURCE	ORGANISM
EST.	Myocosphaerella graminicola.	Myocosphaerella graminicola

REFERENCE	AUTHORS	TITLE	JOURNAL	EDLINE	COMMENT
1 (bases 1 to 587)	Keon, J.P.R., Bailey, A.M. and Hargreaves, J.A.	A group of expressed cDNA sequences from the wheat fungal leaf blotch pathogen, <i>Mycosphaerella graminicola</i> (Septoria tritici)	Fungal Genet. Biol.	29 (2), 118-133 (2000)	20374020
	Contact: Hargreaves JA				

FEATURES
source
1.587
Location/Qualifiers
Seq primer: M13 reverse
Insert length: 1100 Std Error: 0.00
Email: john.hargreaves@bsbc.ac.uk
Fax: +44(0)1275 394281
Tel: +44(0)1275 392181
Long Ashton, Bristol, BS41 9AF, UK

```

/strain="Stilt"-Fraxinella gymniicola"
/db_xref="taxon:54734"
/clone="MGA0121"
/clone_1b="MGA Library"
/note="Vector: pSPORT1; library constructed from cultures
utilising ammonium ions as a source of nitrogen"
144 a 155 c 160 g 128 t

```

Query Match	67.98%	Score 22.4	DB 10	Length 587
Best Local Similarity	81.2%	Pred. 92		
Matches	26	Conservative	0	Mismatches
			6	Indels
				Gaps
				0
2	CCACATGAGTGGCAGGCGTCTGATGATACCG	33		
228	CACACATGATGGCGATGGCAGCTGATGATCGAG	259		

	AV003500	211 bp	mRNA	linear	EST 24-ANG-1999
	Mus musculus	C57BL/6J	kidney Mus	musculus	cDNA clone
	AV003500				
	INITIATION -				
	003500/c				
	US				
	SOLIT 3				

ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS	TITLE JOURNAL COMMENT
U000350E08, mRNA sequence. AY0003500 AY0003500.1 GI:4780350 Est. house mouse. Mus musculus	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 211)	
Carnucci, P., Shibata, K., Osawa, Y., Kono, H., Itoh, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara A., Hayatsu, N., Horii, F., Ishikawa, T., Itoh, M., Iizawa, M., Kawai, J., Kichikuni, N., Kojima, Y., Matsuyama, T., Nitsunuma, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, I., Sogabe, Y., Sugahara Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomihaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Murmatsu, S., Okazaki, Y. and Hayashizaki, Y.	RIKEN Mouse ESTs Unpublished (1999) Contact: Chie Owa

RIKEN Genomic Science Laboratory
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel.: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-reserctc.riken.go.jp
Thermotabilization and thermocycling of thermolabile enzymes by
Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998)
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.reserctc.riken.go.jp>) for
further details

SEQUENCE COUNT	SEQUENCE	SEQUENCE COUNT	SEQUENCE	SEQUENCE COUNT	SEQUENCE
50 a	64 c	50 g	45 t	2	others

Query Match	66.7%;	Score 22;	DB 9;	Length 211;
Best Local Similarity	80.6%;	Fred. No. 85;		
Matches	25;	Conservative	0;	Mismatches
			6;	Indels
			0;	Gaps
				0;
1	GGCCATGATGTCGACGCGCTGCGATAC	31		
83	GGCCATGATGTCGACGCGCTGCGATTC	53		

SPLIT 4
989608/c
CUS
DEFINITION
AQ989608
RfC00171 Photorhabdus luminescens strain W14 M13 library
Photorhabdus luminescens genomic clone PlC00171, DNA sequence.
AQ989608
GSS.
GI:9648202
Photorhabdus luminescens.

```

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?file=6t2-PM3-UMN089-170)
Seq primer: puc 18 forward
High quality sequence stop: 7
High quality sequence start: 76
Location/Qualifiers
1. .118
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="UMN089"
/dev_stage="Adult"
/notes="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT
29 a 22 c 39 g 28 t
ORIGIN
Query Match 65.5%; Score 21.6; DB 10; Length 119;
Best Local Similarity 85.7%; Pred. No. 97;
Matches 24; Conservative % 0; Mismatches 4; Indels 0; Gaps 0;
Oy 3 CACATGAGTGGCAGCGCTCTGTGATA 30'
1 | | | | | | | | | | | | | | | | | | | |
Db 23 CTCATGAGTAGCACAGACGCTTTGTGATA 50'

RESULT 6
LOCUS A2922863/c
DEFINITION A2922863 Sorghum bicolor STcOt Sorghum bicolor genomic DNA
sequence.
ACCESSION A2922863
VERSION A2922863.1 GI:13433084
KEYWORDS GSS.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
1..(bases 1 to 149)
Peterson,D.G., Schulze,S.R., Sclara,E.B., Lee,S.A., Bowers,J.E., Nagel,A., Jiang,N., Tibbits,D.C., Wessler,S.R. and Paterson,A.H. Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery
Genome Res. 12 (5), 795-807 (2002)
21992826
Contact: Peterson DG
Plant Genome Mapping Laboratory
University of Georgia
Room 162, Riverbend Research Bldg., 110 Riverbend Rd., Athens, GA 30602, USA
Tel: 706-583-0167
Fax: 706-583-0160
Email: dgp@arches.uga.edu
Class: Hydroxyapatite-fractionated DNA.
Location/Qualifiers
1. .149
/organism="Sorghum bicolor"
/cultivar="Brx623"
/db_xref="taxon:4558"
/clone_id="Sorghum bicolor STcOt"
/tissue_type="leaves"
/dev_stage="seeding"
/notes="Vector: pGEM-TA-Easy; A Cot analysis was performed

```

for the sorghum genome. Based on the resulting Cot curve, hydroxyapatite chromatography was used to isolate 'highly-repetitive' (HR), 'moderately-repetitive' (MR), and 'single/low-copy' (SL) sequence components from sheared genomic DNA. The three repetition-based DNA components were cloned into E. coli to produce HRCot, MRCot, and SLCot genomic libraries. Blotting and sequencing data indicates that each library is representative of the component from which it was derived. Putative ID listings given for sequences are based on comparison (blastn) with sequences in the NCBI Nr Database. Only the primary match is given (all primary E values are < or = 1.00E-5). In no instance does a 'Cot clone' contain the complete sequence of its putative Nr match."

BASE COUNT 21 a 43 c 46 g 29 t 10 others

Query Match 64.8%; Score 21.4; DB 17; Length 149;
Best Local Similarity 80.6%; Pred. No. 1.3e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GCCACATGAGTGCAGGCGCTGTGTATAC 31
Db 135 GCCTCATGCTGACGACGCGACGTGTGACAC 105

RESULT 7
LOCUS BF676553/c 777 bp mRNA linear EST 21-DEC-2000
DEFINITION 602084468F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4248641 5',
mRNA sequence.
ACCESSION BF676553
VERSION BF676553.1 GI:11950448
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 777)

REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov

COMMENT

Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov

Plate: LICM1067 row: h column: 18
High quality sequence stop: 392.
Location/Qualifiers

FEATURES

Source

1. 777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4248641"
/clone_1b="NIH_MGC_83"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcgatcgcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATCTAGAGCGCGAGCGCCACACAG-dt(30)-Bn-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT

223 a 249 c 117 g 188 t

Query Match 64.2%; Score 21.2; DB 12; Length 777;
Best Local Similarity 88.5%; Pred. No. 3.2e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CACATGAGTGCAGGCGTGTGTGA 28
Db 495 CTCATGAGTGCAGGCGTGTGTGA 470

RESULT 8
LOCUS BE078191 230 bp mRNA linear EST 09-JUN-2000
DEFINITION CM0-BT0615-150200-177-g06 BT0615 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE078191
VERSION BE078191.1 GI:8429758
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 230)

REFERENCE Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LIR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?fl=ft2=CM0-BT0615-150200-177-g06<3=2000-02-15<4=1)
Seq primer: puc 18 forward
High quality sequence start: 32
High quality sequence stop: 118.
Location/Qualifiers

COMMENT

FEATURES

Source

1. 230
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1b="BT0615"
/dev_stage="Adult"

/note="Organ: breast; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 58 a 45 c 65 g 62 t

ORIGIN

Query Match 63.6%; Score 21; DB 10; Length 230;
Best Local Similarity 82.8%; Pred. No. 2.3e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GCCACATGAGTGCAGGCGCTGTGTAT 29
Db 82 GCACATGAGTGCAGGCGCTGTGTAT 110

RESULT 9

AV744190/c 609 bp mRNA linear EST 17-OCT-2000
 LOCUS AV744190 CB Homo sapiens CDNA clone CBUEE04 5', mRNA sequence.
 DEFINITION AV744190
 ACCESSION AV744190
 VERSION AV744190.1 GI:10861771
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 609)
 Zhang, Q., Ye, M., Wu, X., Gu, J., Huang, Q., Zhou, J., Shen, Y., Han, Z.,
 Chen, S., Mao, M. and Chen, Z.
 Homo sapiens CB library cDNA clones
 Unpublished (2000)
 TITLE JOURNAL
 COMMENT Contact: Zhu Chen
 Shanghai Institute of Hematology, Rui-jin Hospital
 197 Rui-jin II Road, Shanghai 200025, P. R. China
 Tel: 86-21-64740490
 Fax: 86-21-64743206
 Email: mbsheims.stn.sh.cn
 This clone is available at Shanghai Hematology Institute in
 Shanghai.
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong.
 Location/Qualifiers
 1..609
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CBUEE04"
 /clone_1lb="CB"
 /tissue_type="cord blood"
 /cell_type="CD34+ hematopoietic stem/progenitor cell"
 /lab_host="BM25.8"
 /note="Vector: pBluescript; Site_1: EcoRI; The insert is
 cloned randomly with the EcoRI digestion"
 165 a 194 c 107 g 141 t 2 others

BASE COUNT
 ORIGIN

Query Match 63.6%; Score 21; DB 10; Length 609;
 Best Local Similarity 82.8%; Pred. No. 3.4e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GCCACATGAGTGGCAAGCGTCTGTGTAT 29
 |||||||||
 DB 538 GCTCATGAGTGGCAAGCGTCTGTGTAT 510

RESULT 10
 BE784735/c 778 bp mRNA linear EST 20-OCT-2000
 LOCUS BE784735
 DEFINITION 601473642P1 NIH_MGC_68 Homo sapiens CDNA clone IMAGE:3876443 5',
 mRNA sequence.
 ACCESSION BE784735
 VERSION BE784735.1 GI:10205933
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 778)
 NIH-MGC http://mgc.ncl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 TITLE JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: DCTD/DMP/gazdar
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LHAM9637 row: d column: 12
 High quality sequence stop: 633.
 Location/Qualifiers
 1..778
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3876443"
 /clone_1lb="NIH_MGC_68"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Lung; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.8 kb. Library constructed by Life
 Technologies."
 BASE COUNT 259 a 237 c 110 g 172 t

ORIGIN

Query Match 63.6%; Score 21; DB 12; Length 778;
 Best Local Similarity 82.8%; Pred. No. 3.8e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GCCACATGAGTGGCAAGCGTCTGTGTAT 29
 |||||||||
 DB 488 GCTCATGAGTGGCTAGACGCTTGTGTAT 460

RESULT 11
 BI296816/c 538 bp mRNA linear EST 19-JUL-2001
 LOCUS BI296816
 DEFINITION UI-R-DKO-cfz-g-01-0-UI.s1 UI-R-DKO Rattus norvegicus CDNA clone
 UI-R-DKO-cfz-g-01-0-UI 3', mRNA sequence.
 ACCESSION BI296816
 VERSION BI296816.1 GI:14961619
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 538)
 Bonaldi, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL 97044477
 MEDLINE
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 normalized rat placenta pool library cDNA library preparation: M.B.
 Soares lab Clone distribution: clones will be available through
 Research Genetics (www.resgen.com)
 Seg primer: M13 Forward
 POLYA=yes.
 Location/Qualifiers
 1..538
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-DKO-cfz-g-01-0-UI"
 /clone_1lb="UI-R-DKO"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DKO

AV737831/c
 LOCUS AV737831 613 bp mRNA linear EST 17-OCT-2000
 DEFINITION AV737831 CB Homo sapiens CDNA clone CBMARH09 5', mRNA sequence.
 ACCESSION AV737831
 VERSION AV737831.1 GI:10855412
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 613)
 Zhang, Q., Ye, M., Wu, X., Gu, J., Huang, Q., Zhou, J., Shen, Y., Han, Z.,
 Chen, S., Mao, M. and Chen, Z.
 Homo sapiens CB library cDNA clones
 Unpublished (2000)
 CONTACT: Zhu Chen
 Shanghai Institute of Hematology, Rui-Jin Hospital
 197 Rui-Jin II Road, Shanghai 200025, P. R. China
 Tel: 86-21-64740490
 Fax: 86-21-64743206
 Email: mbs@shs.sh.cn
 This clone is available at Shanghai Hematology Institute in
 Shanghai.
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong.
 Location/Qualifiers
 1..613
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CBMARH09"
 /clone_1lb="CB"
 /tissue_type="cord blood"
 /cell_type="CD34+ hematopoietic stem/progenitor cell"
 /lab_host="BM25.8"
 /note="Vector: pBluescript; Site_1: EcoRI; The insert is
 cloned randomly with the EcoRI digestion"
 BASE COUNT 168 a 194 c 102 g 146 t
 ORIGIN

Query Match 62.4%; Score 20.6; DB 10; Length 613;
 Best Local Similarity 85.2%; Pred. No. 5e+02;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CACATGAGTGGCAGCGCTGTGTGAT 29
 | ||||| ||||| ||||| ||||| |||||
 DB 536 CTCATGAGTGGCAGCGCTGTGTGAT 510

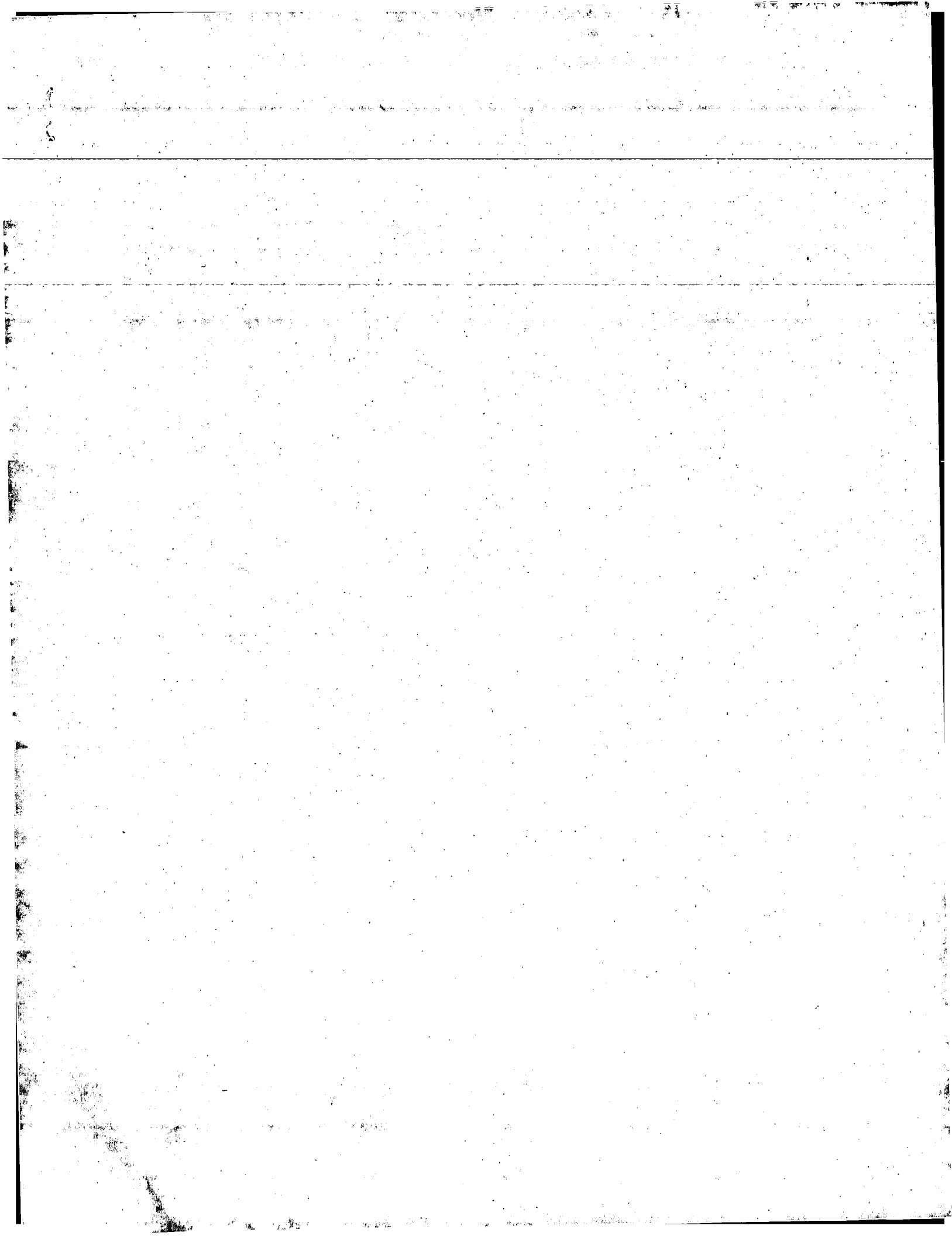
RESULT 15
 AI525664/c
 LOCUS AI525664 777 bp mRNA linear EST 09-AUG-1999
 DEFINITION PT1.3_04_D05.r tumor1 Homo sapiens CDNA 5', mRNA sequence.
 ACCESSION AI525664
 VERSION AI525664.1 GI:4439799
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 777)
 Huang, S.M., Ng, W.L., Farakas, J., He, L., Liang, H.A., Gordon, D., Yu, J.
 and Hood, L.
 Prostate Cancer expression profiling by cDNA sequencing analysis
 Genomics 59 (2), 178-186 (1999)
 99339982
 CONTACT: Guyang Matthew Huang
 Leroy Hood
 University of Washington
 Department of Molecular Biotechnology, Box 357730, University of
 Washington, Seattle, WA 98195
 Tel: 5106280100
 Fax: 5106280108
 Email: huanggm@yahoo.com.

FEATURES
 source Location/Qualifiers
 1..777
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="tumor1"
 /note="Organ: Prostate; Vector: pBluescript; Directional
 cDNA library was constructed using Lambda ZP II kit
 (Stratagene). mRNA was extracted from a frozen prostate
 tumor tissue (Mayo Clinics)."
 BASE COUNT 205 a 189 c 151 g 146 t 86 others
 ORIGIN

Query Match 62.4%; Score 20.6; DB 9; Length 777;
 Best Local Similarity 85.2%; Pred. No. 5.6e+02;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CACATGAGTGGCAGCGCTGTGTGAT 29
 | ||||| ||||| ||||| ||||| |||||
 DB 324 CTCATGAGTGGCAGCGCTGTGTGAT 298

Search completed: July 11, 2003, 02:25:43
 Job time : 600.333 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 : Search time 54.8629 Seconds
(without alignments)
1067.241 Million cell updates/sec

Title: US-10-087-631B-6
Perfect score: 26
Sequence: 1 gcagaagcgtctagccatgagcgcta 26

Scoring table: IDENTITY-NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	26	18	AAAT67193
2	26	100.0	26	19	AAAV59058
3	26	100.0	26	22	AAH25413
4	26	100.0	53	24	AAI40117
5	26	100.0	73	24	AAI40115
6	26	100.0	77	22	AAAI0490
7	26	100.0	155	21	AAZ57775
8	26	100.0	202	12	AAQ14083
9	26	100.0	202	12	AAQ14085

10	26	100.0	202	12	AAQ14086	HCV-N (1-200). He
11	26	100.0	202	12	AAQ14084	HCV-1 (1-202). He
12	26	100.0	242	14	AAQ37774	Cloned HCV 5' non
13	26	100.0	244	19	AAV70449	HCV subtype 1a PCR
14	26	100.0	244	19	AAV70450	HCV subtype 1b PCR
15	26	100.0	244	19	AAV70451	HCV subtype 2c PCR
16	26	100.0	244	24	ABLA6059	Hepatitis C virus
17	26	100.0	244	24	ABLA6060	Hepatitis C virus
18	26	100.0	244	24	ABLA6061	Hepatitis C virus
19	26	100.0	260	21	AAZ57774	Hepatitis C virus
20	26	100.0	279	24	ABA01119	Hepatitis C virus
21	26	100.0	281	17	AAI29118	Hepatitis C virus
22	26	100.0	281	17	AAI29120	Hepatitis C virus
23	26	100.0	282	17	AAI29119	Hepatitis C virus
24	26	100.0	286	19	AAV70444	HCV subtype 1b tar
25	26	100.0	286	24	ABLA6054	Hepatitis C virus
26	26	100.0	289	19	AAV70443	HCV consensus sequ
27	26	100.0	289	19	AAV70445	HCV subtype 2c tar
28	26	100.0	289	24	ABLA6053	Hepatitis C virus
29	26	100.0	289	24	ABLA6055	Hepatitis C virus
30	26	100.0	305	18	AAI87088	HCV amplification
31	26	100.0	305	24	ABN79969	Hepatitis C virus
32	26	100.0	305	24	ABN79970	Hepatitis C virus
33	26	100.0	305	24	ABN79971	Hepatitis C virus
34	26	100.0	305	24	ABN79972	Hepatitis C virus
35	26	100.0	306	15	AAO67079	Hepatitis C virus
36	26	100.0	308	21	AAI75294	Novel hepatitis C
37	26	100.0	312	21	AAI26198	Adapted HCV 5' non
38	26	100.0	314	21	AAI26197	Adapted HCV 5' non
39	26	100.0	323	24	ABK70882	HCV genome 5'UTR s
40	26	100.0	323	24	ABK70883	HCV genome 5'UTR s
41	26	100.0	326	24	ABK70880	HCV genome 5'UTR s
42	26	100.0	327	21	AAI26199	Adapted HCV 5' non
43	26	100.0	327	24	ABK70884	HCV genome 5'UTR s
44	26	100.0	328	18	AAI77074	Hepatitis C virus
45	26	100.0	328	24	ABLA6273	Hepatitis C virus

ALIGNMENTS

RESULT 1	AAAT67193	standard; DNA; 26 BP.
AC	AAAT67193;	
DT	13-FEB-1998	(first entry)
DE	Hepatitis C virus (HCV) RNA amplification primer ST280A.	
KW	Hepatitis C virus; HCV; ST280A; reverse transcription PCR; RT-PCR;	
KW	PCR primer; ss.	
OS	Synthetic.	
PN	EP776981-A2.	
PD	04-JUN-1997.	
PF	21-NOV-1996;	96EP-0118704.
PR	29-NOV-1995;	95US-0007739.
PA	(HOFF) HOFFMANN LA ROCHE & CO AG F.	
PI	Tsang SY;	
XX	WPI: 1997-291296/27.	
DR	Oligonucleotide primers for hepatitis C virus RNA amplification	
PT	by polymerase chain reaction	
XX		

PS Claim 1; Page 11; 16pp; English.

CC This upstream primer ST280A is used in the amplification of the
 CC Hepatitis C virus (HCV) RNA by reverse transcription PCR. This is used
 CC to amplify a 250 base pair product from the 5' untranslated region of
 CC the HCV genome. This can be used to detect HCV in a sample with increased
 CC sensitivity. Amplification of HCV nucleic acid using this primer is up to
 CC 100 times more efficient than amplification with prior art primers.

SQ Sequence 26 BP; 7 A; 6 C; 8 G; 5 T; 0 other;

Query Match 100.0%; Score 26; DB 18; Length 26;
 Best Local Similarity 100.0%; Pred. No. 0.0023;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGCGCTTA 26
 1 GCAGAAAGCGTCTAGCCATGCGCTTA 26

RESULT 2

ID AAV59058 standard; DNA; 26 BP.

AAV59058;

07-JAN-1999 (first entry)

Primer ST280A for HCV fragment.

PCR primer; HCV; nucleic acid amplification; ss.

Synthetic.

Human cytomegalovirus.

Key Location/Qualifiers

modified_base 26

/*tag- a
 /note- "optionally benzylated, methylated, or
 nitrobenzylated"

BP866071-A2.

23-SEP-1998.

12-MAR-1998; 98BP-0104461.

20-MAR-1997; 97US-0041127.

(HOFF) HOFFMANN LA ROCHE & CO AG F.

Will SG, Young KKY.

WPI: 1998-482929/42.

Oligo-nucleotide(s) containing N-substituted nucleotide - useful as
 primers for nucleic acid amplification

Example 6; Page 16; 38pp; English.

This sequence represents a primer for a fragment of HCV, and is an
 example of an oligonucleotide of the invention. The oligonucleotides of
 the invention are of the formula 5'-S1-Nu-3' or 5'-S1-Nu-S2-3', where
 S1 is a sequence of 5-50 nucleotides; S2 is a sequence of 1-3
 nucleotides; and Nu is a nucleotide with a purine or pyrimidine base
 having an exocyclic amino group substituted by a phenyl, phenoxy or optionally
 1-10C-alkyl, alkoxy, optionally substituted phenyl, phenoxy or optionally
 substituted naphthyl. The oligonucleotides are useful as primers for
 nucleic acid amplification, preferably by polymerase chain reaction. Use
 of the modified primers reduces non-specific amplification, especially
 primer dimer formation, with a concomitant increase in the yield of the
 intended target.

SQ Sequence 26 BP; 7 A; 6 C; 8 G; 5 T; 0 other;

Query Match 100.0%; Score 26; DB 19; Length 26;
 Best Local Similarity 100.0%; Pred. No. 0.0023;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGCGCTTA 26
 1 GCAGAAAGCGTCTAGCCATGCGCTTA 26

Db 1 GCAGAAAGCGTCTAGCCATGCGCTTA 26

RESULT 3

ID AAH25413 standard; DNA; 26 BP.

AAH25413;

22-AUG-2001 (first entry)

Forward PCR primer used to amplify a HCV DNA fragment.

Magnetic glass particle; nucleic acid purification; PCR primer; ss.

Hepatitis C virus.

Key Location/Qualifiers

modified_base 26

/*tag- a
 /note- "derivatisation with a p-(t-butyl)benzyl-residue"

WO200137291-A1.

25-MAY-2001.

17-NOV-2000; 2000MO-EP11459.

17-NOV-1999; 98BP-0122853.

12-MAY-2000; 2000BP-0110165.

(HOFF) ROCHE DIAGNOSTICS GMBH.

Weindel K, Riedling M, Geiger A.

WPI: 2001-381247/40.

Novel composition of magnetic glass particles for purification of DNA
 or RNA in automated processes

Example 7; Page 98; 105pp; English.

The specification describes a composition of magnetic glass particles,
 which contain at least one magnetic object with a mean diameter between
 5-500 nm. The composition is useful for the purification of nucleic
 acids. The composition can be used to process large quantities of
 nucleic acid samples, because it does not involve the particles being
 centrifuged or the fluids being drawn through glass fiber filters.
 PCR primers AAH25413-14 were used to amplify HCV DNA fragments. The
 amplified fragment can be purified using the method of the invention.

SQ Sequence 26 BP; 7 A; 6 C; 8 G; 5 T; 0 other;

Query Match 100.0%; Score 26; DB 22; Length 26;
 Best Local Similarity 100.0%; Pred. No. 0.0023;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGCGCTTA 26
 1 GCAGAAAGCGTCTAGCCATGCGCTTA 26

Db 1 GCAGAAAGCGTCTAGCCATGCGCTTA 26

RESULT 4

AA140117 standard; DNA; 53 BP.

```

XX AC AAL40117;
XX XX
XX 13-SEP-2002 (first entry)
XX DT
XX XX
XX Pathogenic microorganism detecting probe SEQ ID No 43.
XX KM Tuberculosis; Mycobacterium; bovis; BCG; africanum; microti; canottii;
XX KW monitoring therapy; pathogenic microorganism; probe; ss.
XX OS
XX Unidentified.
XX PN WO200252043-A1.
XX PD
XX 04-JUL-2002.
XX PF
XX 26-DEC-2001; 2001WO-JP11422.
XX PR
XX 26-DEC-2000; 2000JP-0396222.
XX PR 26-DEC-2000; 2000JP-0396321.
XX PR 29-JUN-2001; 2001JP-0199552.
XX PR 13-SEP-2001; 2001JP-0278920.
XX XX
XX (TAKI ) TAKARA SHUZO CO LTD.
XX PA
XX Shimada M, Hino F, Kato I;
XX PI
XX WPI: 2002-500769/53.
XX DR
XX
XX Detecting pathogenic microorganisms with oligonucleotide probes and
XX PT primers, useful in disease diagnosis and monitoring therapy.
XX XX
XX PS Claim 57; Page 98; 106pp; Japanese.
XX CC The invention relates to a probe containing a 410 or 20 base pair
XX CC sequence, given in the specification. It is capable of detecting the
XX CC tuberculosis bacterial group including Mycobacterium tuberculosis,
XX CC Mycobacterium bovis BCG, Mycobacterium africanum, Mycobacterium microti
XX CC and/or Mycobacterium canottii. The method is useful in disease diagnosis
XX CC and monitoring therapy. This polynucleotide sequence represents a probe
XX CC relating to the detection of pathogenic microorganisms of the invention.
XX SQ
XX Sequence 53 BP; 12 A; 13 C; 17 G; 11 T; 0 other;
XX
Query Match 100.0%; Score 26; DB 24; Length 53;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
DB 2 GCAGAAAGCGTCTAGCCATGGCGTTA 27

```

```

XX XX
XX 26-DEC-2000; 2000JP-0396222.
XX PR 26-DEC-2000; 2000JP-0396321.
XX PR 29-JUN-2001; 2001JP-0199552.
XX PR 13-SEP-2001; 2001JP-0278920.
XX XX
XX (TAKI ) TAKARA SHUZO CO LTD.
XX PA
XX Shimada M, Hino F, Kato I;
XX PI
XX WPI: 2002-500769/53.
XX DR
XX
XX Detecting pathogenic microorganisms with oligonucleotide probes and
XX PT primers, useful in disease diagnosis and monitoring therapy.
XX XX
XX PS Claim 57; Page 97; 106pp; Japanese.
XX CC The invention relates to a probe containing a 410 or 20 base pair
XX CC sequence, given in the specification. It is capable of detecting the
XX CC tuberculosis bacterial group including Mycobacterium tuberculosis,
XX CC Mycobacterium bovis BCG, Mycobacterium africanum, Mycobacterium microti
XX CC and/or Mycobacterium canottii. The method is useful in disease diagnosis
XX CC and monitoring therapy. This polynucleotide sequence represents a PCR
XX CC primer relating to the detection of pathogenic microorganisms of the
XX CC invention.
XX SQ
XX Sequence 73 BP; 18 A; 23 C; 15 G; 17 T; 0 other;
XX
Query Match 100.0%; Score 26; DB 24; Length 73;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
DB 46 GCAGAAAGCGTCTAGCCATGGCGTTA 71

```

```

RESULT 6
AAS10490
ID AAS10490 standard; RNA; 77 BP.
XX AC
XX AAS10490;
XX DT
XX 24-OCT-2001 (first entry)
XX DE
XX HCV 5'-UTR domain II EMSA RNA probe.
XX KW HCV 5'-UTR; minimal IRES; mIRES; internal ribosome entry site; eIF3;
XX KW eukaryotic initiation factor 3; HCV translation initiation; antiviral;
XX KW RNA electrophoretic gel mobility shift assay; EMSA; ss.
XX OS
XX Hepatitis C virus strain Ia M67463.
XX FH
XX Key
XX misc_binding
XX Location/Qualifiers
XX 1..5
XX /tag= a
XX /bound_moiety- "Forms double stranded region with
XX bases 73-77"
XX 8..22
XX /tag= c
XX /note= "Designated as Ila"
XX 23..28
XX /tag= b
XX /bound_moiety- "Forms double stranded region with
XX bases 60-55"
XX 32..50
XX /tag= c
XX /note= "Designated as IId"
XX 55..60
XX /tag= d
XX /bound_moiety- "Forms double stranded region with
XX bases 28-23"
XX 73..77
XX misc_binding
XX

```

[illegible]

CC domain III RNA probe used in a RNA electrophoresis gel mobility
CC shift assay (EMSA). The present sequence is described in an
CC invention relating to a novel compound comprising nucleotide sequences
CC capable of annealing and which is derived from a 5'-untranslated
CC region (UTR) of HCV which is essential for binding of eIF3 (Eukaryotic
CC initiation factor 3). The invention particularly relates to a
CC sub-region of the HCV 5'-UTR referred to as the minimal internal
CC ribosome entry site (IRES), which can be used to identify drugs which
CC inhibit HCV translation initiation. The compounds of the invention may
CC be used to screen for potential HCV antiviral compounds. Assays based
CC on the IRES enable potential antivirals to be screened in a cheaper
CC and easier way. It allows rapid assaying with a small volume of
CC material and are suitable to parallel processing.

Query Match	100.0%	Score 26:	DB 22:	Length 77:
Best Local Similarity	80.8%	Pred. NO.	0.0027:	
Matches	21: Conservative	5: Mismatches	0: Indels	0: Gaps
Qy	1	GCAGAACGCGTCCAGCCATGGCGCTTA	26	
Db	26	GCAGAACCGCTCCAGCCAGCGCGCTTA	51	

	RESULT 7
AA5775/C	
ID	AA5775 standard; DNA, 155 BP.
XX	
AC	AA5775;
XX	
DT	05-APR-2000 (first entry)
XX	
DE	Hepatitis C virus antisense inhibitor oligonucleotide #41.
XX	
KM	Hepatitis C virus; HCV; antisense oligonucleotide; hepatotropic; ss
XX	anti-inflammatory; translation inhibition; HCV infection; virucide.
OS	Hepatitis C virus.
XX	
PN	US6001990-A.
XX	
PD	14-DEC-1999.
XX	
PE	07-JUN-1995; 95US-0474700.
XX	
FR	10-MAY-1994; 94US-0240382.
XX	

PA (GEHO) GEN HOSPITAL CORP.
XX
XX
PI Moradpour D, Wands JR, Wakita T
XX
XX WPI; 2000-104900/09.
XX

PT	Antisense oligonucleotide to Hepatitis C virus RNA, useful for treating
PT	Hepatitis C virus infections -
PS	
XX	Claim 30; Column 33; 31pp; English.
XX	

ins sequence is an antisense oligonucleotide that hybridizes to Hepatitis C virus (HCV) RNA, under physiological conditions. The invention relates to HCV antisense oligonucleotides, and also for a vector comprising a nucleotide sequence which is transcribed in an animal cell to generate an antisense oligonucleotide. The oligonucleotides have virucide, hepatotropic and anti-inflammatory activity, and are useful for treating HCV infection by inhibiting translation of type I-V HCV RNA. Hepatitis C virus is a positive strand RNA virus, and is the major causative agent of post-transfusion hepatitis. Persistent HCV infection can lead to chronic hepatitis, cirrhosis, and hepatocellular carcinoma.

Query Match	100.0%	Score 26	DB 21	Length 155
Best Local Similarity	100.0%	Pred. No. 0.003		
Matches 26; Conservative	0	Mismatches	0	104.7%

GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
GCAGAAAGCGTCTAGCCATGGCGTTA 63

RESULT	8
AAQ14083	
AAQ14083	standard; DNA; 202 BP

03-JAN-1992 (first entry)
HCV-T (1-202).
Epitope; PCR; diagnosis; ss
Hepatitis C virus.

W091147/9-A.	
03-OCT-1991.	
28-MAR-1991.	91WO-JP00405.
09-NOV-1990.	90JP-0305795.
28-MAR-1990.	90JP-0080185.
13-JUN-1990.	90JP-0154280.
14-JUN-1990.	90JP-0133979.

(MAIN) MITSUBI TOATSU CHEM INC.
Takada T, Enomoto N, Date T, Nakao T;
WPI; 1991-310579/42.

Disclosures: Fig 1(I): 79pp; Japanese.

encode epitopes from structural, non-structural and 5' untranslated domains of hepatitis C virus. The sequences are used for accurate and simple diagnosis and typing of HCV infection, using PCR

CC amplification techniques.

XX Sequence 202 BP; 39 A; 61 C; 61 G; 41 T; 0 other;

XX Query Match 100.0%; Score 26; DB 12; Length 202;

XX Best Local Similarity 100.0%; Pred. No. 0.0032;

XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 |||||||
 DB 25 GCAGAAAGCGTCTAGCCATGGCGTTA 50

RESULT 9

AA014085 ID AA014085 standard; DNA; 202 BP.

AC AA014085;

DT 03-JAN-1992 (first entry)

DE HCV-KU (1-202).

KM Epitope; PCR; diagnosis; ss.

OS Hepatitis C virus.

PN WO9114779-A.

PD 03-OCT-1991.

PF 28-MAR-1991; 91WO-JP00405.

PR 09-NOV-1990; 90JP-0305795.

PR 28-MAR-1990; 90JP-0080185.

PR 13-JUN-1990; 90JP-0154230.

PR 14-JUN-1990; 90JP-0153979.

PA (MITK) MITSUI TOATSU CHEM INC.

PI Takada T, Enomoto N, Date T, Nakao T;

DR WPI; 1991-310579/42.

PS Disclosure: Fig 1(1); 79pp; Japanese.

CC The nucleotide sequences represented in AA014076-86 and AA014767-71

CC encode epitopes from structural, non-structural and 5' untranslated

CC domains of hepatitis C virus. The sequences are used for accurate

CC and simple diagnosis and typing of HCV infection, using PCR

CC amplification techniques.

XX Sequence 202 BP; 40 A; 61 C; 58 G; 43 T; 0 other;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 |||||||
 DB 25 GCAGAAAGCGTCTAGCCATGGCGTTA 50

DB 25 GCAGAAAGCGTCTAGCCATGGCGTTA 50

RESULT 10

AA014086 ID AA014086 standard; DNA; 202 BP.

AC AA014086;

DT 03-JAN-1992 (first entry)

XX HCV-N (1-200).

XX Epitope; PCR; diagnosis; ss.

XX Hepatitis C virus.

PN WO9114779-A.

PD 03-OCT-1991.

PF 28-MAR-1991; 91WO-JP00405.

PR 09-NOV-1990; 90JP-0305795.

PR 28-MAR-1990; 90JP-0080185.

PR 13-JUN-1990; 90JP-0154230.

PR 14-JUN-1990; 90JP-0153979.

PA (MITK) MITSUI TOATSU CHEM INC.

PI Takada T, Enomoto N, Date T, Nakao T;

DR WPI; 1991-310579/42.

PS Disclosure: Fig 1(1); 79pp; Japanese.

CC The nucleotide sequences represented in AA014076-86 and AA014767-71

CC encode epitopes from structural, non-structural and 5' untranslated

CC domains of hepatitis C virus. The sequences are used for accurate

CC and simple diagnosis and typing of HCV infection, using PCR

CC amplification techniques.

XX Sequence 202 BP; 42 A; 60 C; 56 G; 44 T; 0 other;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 |||||||
 DB 25 GCAGAAAGCGTCTAGCCATGGCGTTA 50

RESULT 11

AA014084 ID AA014084 standard; DNA; 202 BP.

AC AA014084;

DT 03-JAN-1992 (first entry)

DE HCV-I (1-202).

KM Epitope; PCR; diagnosis; ss.

OS Hepatitis C virus.

PN WO9114779-A.

PD 03-OCT-1991.

PF 28-MAR-1991; 91WO-JP00405.

PR 09-NOV-1990; 90JP-0305795.

PR 28-MAR-1990; 90JP-0080185.

PR 13-JUN-1990; 90JP-0154230.

PR 14-JUN-1990; 90JP-0153979.

PA (MITK) MITSUI TOATSU CHEM INC.

PI Takada T, Enomoto N, Date T, Nakao T;
 XX
 DR WPI: 1991-310579/42.
 XX
 PT New nucleotide sequences encoding HCV epitope(s) - for diagnosis
 XX of hepatitis C virus infection via polymerase chain reaction
 PS Disclosure: Fig 1(1); 79pp; Japanese.
 CC The nucleotide sequences represented in AA014076-86 and AA014767-71
 CC encode epitopes from structural, non-structural and 5' untranslated
 CC domain of hepatitis C virus. The sequences are used for accurate
 CC and simple diagnosis and typing of HCV infection, using PCR
 CC amplification techniques.
 XX
 SQ Sequence 202 BP; 41 A; 60 C; 60 G; 41 T; 0 other;
 Query Match 100.0%; Score 26; DB 12; Length 202;
 Best Local Similarity 100.0%; Pred. No. 0.0032;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
 25 GCAGAAAGCGCTAGCCATGGCGTTA 50
 RESULT 12
 ID AA037774 standard; cDNA; 242 BP.
 AC AA037774;
 XX
 DT 30-JUN-1993 (first entry)
 XX
 DE Cloned HCV 5' non coding region from pGHCV1A.
 KW Hepatitis C virus; probe; hepatocellular necrosis; hepatocellular;
 KW carcinoma; diagnosis; therapy; ss.
 OS Hepatitis C virus.
 PN EP531974-A.
 XX
 PD 17-MAR-1993.
 XX
 PF 09-SEP-1992; 92EP-0115426.
 XX
 PR 12-SEP-1991; 91US-0758662.
 XX
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 XX
 PI Hu K, Vierling JM;
 XX
 DR WPI: 1993-087007/11.
 XX
 PT Detection of hepatitis C virus (HCV) RNA - using nucleic acid
 XX probes derived from the 5'-non-coding region of the HCV genome
 PS Claim 1; Fig 4; 26pp; English.
 XX
 CC To obtain HCV cDNA nucleotide sequences from the 5' non-coding
 CC region a pair of oligonucleotides based on the reported sequence of
 CC HCV-J1 were used as primers for HCV PCR. HCV RNA was isolated from
 CC serum of a putatively infected individual. RNA reverse
 CC transcription PCR was performed and a specific PCR prod. identified.
 CC The prod. was used to transform E. coli DH5 alpha to obtain pGHCV1A
 CC contg. a 242 bp insertion from the HCV 5' non-coding region. This
 CC probe is highly specific and sensitive for HCV RNA. The probe can
 CC be used to quantitatively detect the amt. of HCV in samples, to
 CC analyse the molecular forms of HCV RNA during evolution of the
 CC disease, to localise HCV in hepatic and/or extrahepatic tissues
 CC and to study the relationship between HCV infection, hepatocellular
 CC necrosis and hepatocellular carcinoma. The probe can be used to

CC diagnose HCV infection, to prepare blood free of HCV and to monitor
 CC anti-HCV therapy.
 XX
 SQ Sequence 242 BP; 51 A; 74 C; 67 G; 50 T; 0 other;
 Query Match 100.0%; Score 26; DB 14; Length 242;
 Best Local Similarity 100.0%; Pred. No. 0.0033;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
 45 GCAGAAAGCGCTAGCCATGGCGTTA 70
 RESULT 13
 ID AA070449 standard; DNA; 244 BP.
 XX
 AC AA070449;
 XX
 DT 08-APR-1999 (first entry)
 XX
 DE HCV subtype 1a PCR fragment.
 XX
 KW Nucleic acid detection; nucleic acid characterisation; hybridisation;
 KW infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.
 OS Hepatitis C virus.
 XX
 PN WC9850403-A1.
 XX
 PD 12-NOV-1998.
 XX
 PF 05-MAY-1998; 98WO-US03194.
 XX
 PR 03-MAR-1998; 98US-0034205.
 XX
 PR 05-MAY-1997; 97US-0851588.
 XX
 PR 19-SEP-1997; 97US-0934097.
 XX
 PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
 XX
 PI Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;
 PI Lyamatchev VI, Neri BP, Prudent JR;
 XX
 DR WPI: 1998-610317/51.
 XX
 PT detection and characterisation of nucleic acid sequences - by mixing
 XX a folded target and one or more probes to form a probe/folded target
 PT complex and detecting and characterising the complexes
 PS Example 3; Page 169; 279pp; English.
 XX
 CC The invention relates to methods and compositions of detection and
 CC characterisation of nucleic acid sequences and sequence changes. One
 CC method of detection and characterisation comprises: (a) providing: (i) a
 CC folded target having a DNA sequence comprising at least 1 double
 CC stranded region and at least 1 single stranded region; and (ii) at least
 CC 1 probe complementary to at least a portion of the folded target; and
 CC (b) mixing the target and probes so that the probe hybridises to form a
 CC of structure formation in nucleic acid targets; for analysing folded
 CC nucleic acids targets; and for analysis of nucleic acid structures. The
 CC methods can be used for the detection and characterisation of nucleic
 CC acid sequences indicative of an infection, the presence of pathogenic nucleic acid
 CC sequences indicative of an infection, the presence of variants or alleles
 CC of mammalian genes associated with disease and cancers, and the
 CC identification of the source of nucleic acids found in forensic samples,
 CC as well as in paternity determinations. The methods allow simultaneous
 CC analysis of both strands (e.g. the sense and antisense strands) and are
 CC ideal for high-level multiplexing. The products produced are amenable to
 CC qualitative, quantitative and positional analysis. The methods may be
 CC performed in solution or in the solid phase (e.g. on a solid support).
 CC The methods are powerful in that they allow for analysis of longer

CC fragments of nucleic acid than current methodologies. Sequences
 CC AAV70449-52 represent a hepatitis C virus (HCV) subtype sequences
 CC produced by PCR. These PCR products can be used in hybridisation analysis
 CC using multiple capture probes for HCV genotyping.

CC Sequence 244 BP; 46 A; 67 C; 80 G; 51 T; 0 other;

Query Match 100.0%; Score 26; DB 19; Length 244;

Best Local Similarity 100.0%; Pred. No. 0.0033; Mismatches 0; Indels 0; Gaps 0;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GCAGAAAGCGCTAGCCATGCGCTTA 26

RESULT 14
 ID AAV70450 standard; DNA; 244 BP.

AC AAV70450;

DT 08-APR-1999 (first entry)

DE HCV subtype 1b PCR fragment.

KM Nucleic acid detection; nucleic acid characterisation; hybridisation;
 KM infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.
 OS Hepatitis C virus.

PN MO9850403-A1.

PD 12-NOV-1998.

PF 05-MAY-1998; 98MO-US03194.

PR 03-MAR-1998; 98US-0034205.

PR 05-MAY-1997; 97US-0851588.

PR 19-SEP-1997; 97US-0934097.

PA (THIR-) THIRD WAVE TECHNOLOGIES INC.

PI Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;

PI Lyamlichev VI, Neri BP, Prudent JR;

DR WPI; 1998-610317/51.

PT Detection and characterisation of nucleic acid sequences - by mixing
 PT a folded target and one or more probes to form a probe/folded target
 PT complex and detecting and characterising the complexes

Example 3; Page 169; 279pp; English.

CC The invention relates to methods and compositions of detection and
 CC characterisation of nucleic acid sequences and sequence changes. One
 CC method of detection and characterisation comprises: (a) providing: (1) a
 CC folded target having a DNA sequence comprising at least 1 double
 CC stranded region and at least 1 single stranded region; and (ii) at least
 CC 1 probe complementary to at least a portion of the folded target; and
 CC (b) mixing the target and probes so that the probe hybridises to form a
 CC probe/folded target complex. Also provided are methods for determination
 CC of structure formation in nucleic acid targets; for analysing folded
 CC nucleic acids targets; and for analysis of nucleic acid structures. The
 CC methods can be used for the detection and characterisation of nucleic
 CC acid sequences to detect the presence of pathogenic nucleic acid
 CC sequences indicative of an infection, the presence of variants or alleles
 CC of mammalian genes associated with disease and cancers, and the
 CC identification of the source of nucleic acids found in forensic samples,
 CC as well as in paternity determinations. The methods allow simultaneous
 CC analysis of both strands (e.g. the sense and antisense strands) and are
 CC ideal for high-level multiplexing. The products produced are amenable to
 CC qualitative, quantitative and positional analysis. The methods may be

CC performed in solution or in the solid phase (e.g. on a solid support).
 CC The methods are powerful in that they allow for analysis of longer
 CC fragments of nucleic acid than current methodologies. Sequences
 CC AAV70449-52 represent a hepatitis C virus (HCV) subtype sequences
 CC produced by PCR. These PCR products can be used in hybridisation analysis
 CC using multiple capture probes for HCV genotyping.

CC Sequence 244 BP; 44 A; 67 C; 81 G; 52 T; 0 other;

Query Match 100.0%; Score 26; DB 19; Length 244;

Best Local Similarity 100.0%; Pred. No. 0.0033; Mismatches 0; Indels 0; Gaps 0;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GCAGAAAGCGCTAGCCATGCGCTTA 26

RESULT 15
 ID AAV70451 standard; DNA; 244 BP.

AC AAV70451;

DT 08-APR-1999 (first entry)

DE HCV subtype 2c PCR fragment.

KM Nucleic acid detection; nucleic acid characterisation; hybridisation;
 KM infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.
 OS Hepatitis C virus.

PN MO9850403-A1.

PD 12-NOV-1998.

PF 05-MAY-1998; 98MO-US03194.

PR 03-MAR-1998; 98US-0034205.

PR 05-MAY-1997; 97US-0851588.

PR 19-SEP-1997; 97US-0934097.

PA (THIR-) THIRD WAVE TECHNOLOGIES INC.

PI Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;

PI Lyamlichev VI, Neri BP, Prudent JR;

DR WPI; 1998-610317/51.

PT Detection and characterisation of nucleic acid sequences - by mixing
 PT a folded target and one or more probes to form a probe/folded target
 PT complex and detecting and characterising the complexes

Example 3; Page 169; 279pp; English.

CC The invention relates to methods and compositions of detection and
 CC characterisation of nucleic acid sequences and sequence changes. One
 CC method of detection and characterisation comprises: (a) providing: (1) a
 CC folded target having a DNA sequence comprising at least 1 double
 CC stranded region and at least 1 single stranded region; and (ii) at least
 CC 1 probe complementary to at least a portion of the folded target; and
 CC (b) mixing the target and probes so that the probe hybridises to form a
 CC probe/folded target complex. Also provided are methods for determination
 CC of structure formation in nucleic acid targets; for analysing folded
 CC nucleic acids targets; and for analysis of nucleic acid structures. The
 CC methods can be used for the detection and characterisation of nucleic
 CC acid sequences to detect the presence of pathogenic nucleic acid
 CC sequences indicative of an infection, the presence of variants or alleles
 CC of mammalian genes associated with disease and cancers, and the
 CC identification of the source of nucleic acids found in forensic samples,
 CC as well as in paternity determinations. The methods allow simultaneous
 CC analysis of both strands (e.g. the sense and antisense strands) and are

CC ideal for high-level multiplexing. The products produced are amenable to
 CC qualitative, quantitative and positional analysis. The methods may be
 CC performed in solution or in the solid phase (e.g. on a solid support).
 CC The methods are powerful in that they allow for analysis of longer
 CC fragments of nucleic acid than current methodologies. Sequences
 CC AAV70449-52 represent a hepatitis C virus (HCV) subtype sequences
 CC produced by PCR. These PCR products can be used in hybridisation analysis
 CC using multiple capture probes for HCV genotyping.

XX Sequence 244 BP; 46 A; 68 C; 78 G; 52 T; 0 other;

Query Match 100.0%; Score 26; DB 19; Length 244;
 Best Local Similarity 100.0%; Pred. No. 0.0033;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCCATGCGCTTA 26
 ||||||||||||||||||||
 Db 1 GCAGAAAGCGCTAGCCATGCGCTTA 26

Search completed: July 10, 2003, 19:52:41
 Job time : 56.8629 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 11.627 Seconds
(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631B-6

Sequence: 1 gcagaaagcgtctacgcgcgtta 26

Scoring table: IDENTITY_MUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PTOUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	100.0	26	US-08-738-928-1	Sequence 1, Appl
2	26	100.0	26	US-09-039-866-3	Sequence 3, Appl
3	26	100.0	155	US-08-474-700B-41	Sequence 41, Appl
4	26	100.0	194	US-08-634-797-44	Sequence 44, Appl
5	26	100.0	194	US-08-634-797-45	Sequence 45, Appl
6	26	100.0	194	US-08-634-797-46	Sequence 46, Appl
7	26	100.0	194	US-08-634-797-47	Sequence 47, Appl
8	26	100.0	194	US-08-634-797-49	Sequence 49, Appl
9	26	100.0	194	US-08-634-797-50	Sequence 50, Appl
10	26	100.0	194	US-08-634-797-51	Sequence 51, Appl
11	26	100.0	194	US-08-634-797-52	Sequence 52, Appl
12	26	100.0	244	US-08-335-595-1	Sequence 1, Appl
13	26	100.0	244	US-09-034-205-26	Sequence 26, Appl
14	26	100.0	244	US-09-034-205-27	Sequence 27, Appl
15	26	100.0	244	US-09-034-205-28	Sequence 28, Appl
16	26	100.0	244	US-08-934-097A-26	Sequence 26, Appl
17	26	100.0	244	US-08-934-097A-27	Sequence 27, Appl
18	26	100.0	244	US-08-934-097A-28	Sequence 28, Appl
19	26	100.0	244	US-08-851-588-26	Sequence 26, Appl
20	26	100.0	244	US-08-851-588-27	Sequence 27, Appl
21	26	100.0	244	US-08-851-588-28	Sequence 28, Appl
22	26	100.0	244	US-09-677-218B-26	Sequence 26, Appl
23	26	100.0	244	US-09-677-218B-27	Sequence 27, Appl
24	26	100.0	244	US-09-677-218B-28	Sequence 28, Appl
25	26	100.0	244	US-09-677-192-26	Sequence 26, Appl
26	26	100.0	244	US-09-677-192-27	Sequence 27, Appl
27	26	100.0	244	US-09-677-192-28	Sequence 28, Appl

C	28	26	100.0	260	3	US-08-474-700B-40	Sequence 40, Appl
	29	26	100.0	281	2	US-08-757-653-123	Sequence 123, App
	30	26	100.0	281	2	US-08-757-653-125	Sequence 125, App
	31	26	100.0	281	2	US-08-757-653-128	Sequence 128, App
	32	26	100.0	281	2	US-08-757-653-129	Sequence 129, App
	33	26	100.0	281	2	US-08-757-653-131	Sequence 131, App
	34	26	100.0	281	4	US-08-520-946-125	Sequence 125, App
	35	26	100.0	281	4	US-08-520-946-128	Sequence 128, App
	36	26	100.0	281	4	US-08-520-946-129	Sequence 129, App
	37	26	100.0	281	4	US-08-520-946-131	Sequence 131, App
	38	26	100.0	281	4	US-08-520-946-132	Sequence 132, App
	39	26	100.0	282	2	US-08-757-653-134	Sequence 134, App
	40	26	100.0	282	2	US-08-757-653-130	Sequence 130, App
	41	26	100.0	282	4	US-08-520-946-124	Sequence 124, App
	42	26	100.0	282	4	US-08-520-946-130	Sequence 130, App
	43	26	100.0	286	4	US-09-034-205-21	Sequence 21, Appl
	44	26	100.0	286	4	US-08-934-097A-21	Sequence 21, Appl
	45	26	100.0	286	4	US-08-851-588-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1

US-08-738-928-1

Sequence 1, Application US/08738928

Patent No. 5837442

GENERAL INFORMATION:

APPLICANT: Tsang, Sue Y.

TITLE OF INVENTION: Oligonucleotide primers for Amplifying

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: U.S.A.

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/738,928

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Petry, Douglas A.

REGISTRATION NUMBER: 35,321

REFERENCE/DOCKET NUMBER: 9263

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 814-2974

TELEFAX: (510) 814-2974

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 26 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-738-928-1

Query Match

Best Local Similarity 100.0%; Score 26; DB 2; Length 26;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTACCGCGGCTTA 26

DB 1 GCAGAAAGCGTCTACCGCGGCTTA 26

RESULT 2
US-09-039-866-3Sequence 3, Application US/09039866
Patent No. 6001611

GENERAL INFORMATION:

APPLICANT: Will, Stephen G.

TITLE OF INVENTION: MODIFIED NUCLEIC ACID AMPLIFICATION

TITLE OF INVENTION: PRIMERS

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Roche Molecular Systems

STREET: 1080 U.S. Highway 202

CITY: Branchburg

STATE: New Jersey

COUNTRY: United States

ZIP: 08876

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/039,866

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Petry, Douglas A.

REGISTRATION NUMBER: 35,321

REFERENCE/DOCKET NUMBER: 1023P

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 26 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-039-866-3

Query Match

Best Local Similarity 100.0%; Score 26; DB 3; Length 26;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/240,382

FILING DATE: 10 May 1994

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.

REGISTRATION NUMBER: 34,819

REFERENCE/DOCKET NUMBER: 00786/279001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 41:

SEQUENCE CHARACTERISTICS:

LENGTH: 155 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-474-700B-41

Query Match

Best Local Similarity 100.0%; Score 26; DB 3; Length 155;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 4
US-08-634-797-44

Sequence 44, Application US/08634797

Patent No. 5851759

GENERAL INFORMATION:

APPLICANT: WEINER, AMY J.

TITLE OF INVENTION: HETEROIDPLEX TRACKING ASSAY (HTA) FOR

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation

STREET: 4560 Horton Street - R440

CITY: Emeryville

STATE: California

COUNTRY: USA

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/634,797

FILING DATE: 19-APR-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hardin, Ailsa A.

REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 1226.001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 601-3274

TELEFAX: (510) 655-3542

TELEX: N/A

INFORMATION FOR SEQ ID NO: 44:

SEQUENCE CHARACTERISTICS:

LENGTH: 194 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-634-797-44

Query Match

Best Local Similarity 100.0%; Score 26; DB 2; Length 194;

Matches 100.0%; Pred. No. 0.00086;

Db

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

1 GCAGAAAGCGCTAGCCATGGCGTTA 26

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
DB 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26

RESULT 5

US-08-634-797-45
Sequence 45, Application US/08634797
Patent No. 5851759

GENERAL INFORMATION:
APPLICANT: WEINER, AMY J.
TITLE OF INVENTION: HETERODUPLEX TRACKING ASSAY (HTA) FOR
GENOTYPING HCV
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street - R440
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/634,797
APPLICATION NUMBER: US/08/634,797
FILING DATE: 19-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hardin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 1226.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-3274
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-634-797-45

Query Match 100.0%; Score 26; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
DB 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26

RESULT 6

US-08-634-797-46
Sequence 46, Application US/08634797
Patent No. 5851759

GENERAL INFORMATION:
APPLICANT: WEINER, AMY J.
TITLE OF INVENTION: HETERODUPLEX TRACKING ASSAY (HTA) FOR
GENOTYPING HCV
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street - R440
CITY: Emeryville
STATE: California
COUNTRY: USA

ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,797
FILING DATE: 19-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hardin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 1226.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-3274
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-634-797-46

Query Match 100.0%; Score 26; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
DB 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26

RESULT 7

US-08-634-797-47
Sequence 47, Application US/08634797
Patent No. 5851759

GENERAL INFORMATION:
APPLICANT: WEINER, AMY J.
TITLE OF INVENTION: HETERODUPLEX TRACKING ASSAY (HTA) FOR
GENOTYPING HCV
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street - R440
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/634,797
APPLICATION NUMBER: US/08/634,797
FILING DATE: 19-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hardin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 1226.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-3274
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-634-797-47

Query Match
Best Local Similarity 100.0%; Score 26; DB 2; Length 194;
Matches 26; Conservative 0; Pred. No. 0.00086;
Mismatched 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCATGCGCTTA 26
Db 1 GCAGAAAGCGCTAGCATGCGCTTA 26

RESULT 8
US-08-634-797-49

Sequence 49, Application US/08634797
Patent No. 5851759

GENERAL INFORMATION:

APPLICANT: WEINER, AMY J.

TITLE OF INVENTION: HETEROIDPLEX TRACKING ASSAY (HTA) FOR

TITLE OF INVENTION: GENOTYPING HCV

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation

STREET: 4560 Horton Street - R440

CITY: Emeryville

STATE: California

COUNTRY: USA

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/634,797

FILING DATE: 19-APR-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hardin, Alisa A.

REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 1226.001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 601-3274

TELEFAX: (510) 655-3542

TELEX: N/A

INFORMATION FOR SEQ ID NO: 49:

SEQUENCE CHARACTERISTICS:

LENGTH: 194 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-634-797-49

Query Match

Best Local Similarity 100.0%; Score 26; DB 2; Length 194;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCATGCGCTTA 26
Db 1 GCAGAAAGCGCTAGCATGCGCTTA 26

RESULT 9
US-08-634-797-50

Sequence 50, Application US/08634797
Patent No. 5851759

GENERAL INFORMATION:

APPLICANT: WEINER, AMY J.

TITLE OF INVENTION: HETEROIDPLEX TRACKING ASSAY (HTA) FOR

TITLE OF INVENTION: GENOTYPING HCV

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation

STREET: 4560 Horton Street - R440

CITY: Emeryville

STATE: California

COUNTRY: USA

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/634,797

FILING DATE: 19-APR-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hardin, Alisa A.

REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 1226.001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 601-3274

TELEFAX: (510) 655-3542

TELEX: N/A

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:

LENGTH: 194 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-634-797-50

Query Match

Best Local Similarity 100.0%; Score 26; DB 2; Length 194;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCATGCGCTTA 26
Db 1 GCAGAAAGCGCTAGCATGCGCTTA 26

RESULT 10
US-08-634-797-51

Sequence 51, Application US/08634797
Patent No. 5851759

GENERAL INFORMATION:

APPLICANT: WEINER, AMY J.

TITLE OF INVENTION: HETEROIDPLEX TRACKING ASSAY (HTA) FOR

TITLE OF INVENTION: GENOTYPING HCV

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation

STREET: 4560 Horton Street - R440

CITY: Emeryville

STATE: California

COUNTRY: USA

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/634,797

FILING DATE: 19-APR-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hardin, Alisa A.

REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 1226.001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 601-3274
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-634-797-51

Query Match 100.0%; Score 26; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCATGGCGTTA 26
|||||
DB 1 GCAGAAAGCGCTAGCATGGCGTTA 26

RESULT 11
US-08-634-797-52
Sequence 52, Application US/08634797
Patent No. 5851759
GENERAL INFORMATION:
APPLICANT: WEINER, AMY J.
TITLE OF INVENTION: HETERO DUPLEX TRACKING ASSAY (HTA) FOR
TITLE OF INVENTION: GENOTYPING HCY
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street - RA40
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,797
FILING DATE: 19-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 1226.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-3274
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-634-797-52

Query Match 100.0%; Score 26; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCATGGCGTTA 26
|||||
DB 1 GCAGAAAGCGCTAGCATGGCGTTA 26

RESULT 12

US-08-335-595-1
Sequence 1, Application US/08335595
Patent No. 5914228
GENERAL INFORMATION:
APPLICANT: VIERLING, JOHN M
APPLICANT: HU, KE-QIN
TITLE OF INVENTION: DIRECT DETECTION OF HEPATITIS C VIRUS
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 611 WEST 6TH STREET
CITY: LOS ANGELES
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/335,595
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/175,473
FILING DATE:
APPLICATION NUMBER: US/07/758,862
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SCHNEIDER, CAROL A
REGISTRATION NUMBER: 34,923
REFERENCE/DOCKET NUMBER: 194/285
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-489-1600
TELEFAX: 213-955-0440
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 242 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-335-595-1

Query Match 100.0%; Score 26; DB 2; Length 242;
Best Local Similarity 100.0%; Pred. No. 0.00089;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCATGGCGTTA 26
|||||
DB 45 GCAGAAAGCGCTAGCATGGCGTTA 70

RESULT 13
US-09-034-205-26
Sequence 26, Application US/09034205
Patent No. 6194149
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance P.
APPLICANT: Neill, Bruce P.
TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING
STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,205
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ. ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-034-205-26

Query Match
Best Local Similarity 100.0%; Score 26; DB 4; Length 244;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 GCAGAAAGCGCTAGCCATGGCGTTA 26
1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 14
US-09-034-205-27
Sequence 27, Application US/09034205
Patent No. 6194149
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Nerl, Bruce P.
TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING
STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,205
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ. ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-034-205-27

Query Match
Best Local Similarity 100.0%; Score 26; DB 4; Length 244;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 GCAGAAAGCGCTAGCCATGGCGTTA 26
1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 15
US-09-034-205-28
Sequence 28, Application US/09034205
Patent No. 6194149
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Nerl, Bruce P.
TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING
STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,205
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ. ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-034-205-28

Query Match
Best Local Similarity 100.0%; Score 26; DB 4; Length 244;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 GCAGAAAGCGCTAGCCATGGCGTTA 26
1 GCAGAAAGCGCTAGCCATGGCGTTA 26

Search completed: July 10, 2003, 20:27:39
Job time : 16.627 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:53:16 ; Search time 71.164 Seconds
(without alignments)
576.891 Million cell updates/sec

Title: US-10-087-631B-6
Perfect score: 26
Sequence: 1 gcagaaagcgtctagccatgacgtgta 26

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/PCRUS_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	26	100.0	26	9 US-10-087-631B-6	Sequence 6, Appli
2	26	100.0	30	12 US-10-029-907-17	Sequence 17, Appl
3	26	100.0	241	9 US-10-087-631B-10	Sequence 10, Appl
4	26	100.0	241	9 US-10-087-631B-13	Sequence 13, Appl
5	26	100.0	242	9 US-10-087-631B-15	Sequence 15, Appl
6	26	100.0	244	10 US-09-825-574-26	Sequence 26, Appl
7	26	100.0	244	10 US-09-825-574-27	Sequence 27, Appl
8	26	100.0	244	10 US-09-825-574-28	Sequence 28, Appl
9	26	100.0	281	9 US-09-940-925A-123	Sequence 123, App
10	26	100.0	281	9 US-09-940-925A-125	Sequence 125, App
11	26	100.0	281	9 US-09-940-925A-128	Sequence 128, App
12	26	100.0	281	9 US-09-940-925A-129	Sequence 129, App
13	26	100.0	281	9 US-09-941-193A-131	Sequence 131, App
14	26	100.0	281	9 US-09-941-193A-133	Sequence 133, App
15	26	100.0	281	9 US-09-941-193A-125	Sequence 125, App
16	26	100.0	281	9 US-09-941-193A-128	Sequence 128, App
17	26	100.0	281	9 US-09-941-193A-129	Sequence 129, App
18	26	100.0	281	9 US-09-941-193A-131	Sequence 131, App
19	26	100.0	282	9 US-09-940-925A-124	Sequence 124, App

c	20	26	100.0	282	9	US-09-940-925A-130	Sequence 130, App
	21	26	100.0	282	9	US-09-941-193A-124	Sequence 124, App
c	22	26	100.0	282	9	US-09-941-193A-130	Sequence 130, App
	23	26	100.0	286	10	US-09-825-574-21	Sequence 21, Appl
	24	26	100.0	289	10	US-09-825-574-20	Sequence 20, Appl
	25	26	100.0	289	10	US-09-825-574-22	Sequence 22, Appl
	26	26	100.0	298	10	US-09-345-761-7	Sequence 7, Appl
	27	26	100.0	315	10	US-09-345-761-6	Sequence 6, Appl
	28	26	100.0	341	9	US-09-814-357-3	Sequence 3, Appl
	29	26	100.0	341	9	US-10-259-275-35	Sequence 35, Appl
	30	26	100.0	341	10	US-09-814-292-44	Sequence 44, Appl
	31	26	100.0	347	1	US-10-132-295-1	Sequence 1, Appl
	32	26	100.0	366	9	US-09-992-160-48	Sequence 48, Appl
	33	26	100.0	366	9	US-09-740-332-9701	Sequence 9701, Ap
	34	26	100.0	366	10	US-09-877-526A-48	Sequence 48, Appl
	35	26	100.0	375	9	US-10-087-631B-14	Sequence 14, Appl
	36	26	100.0	386	9	US-09-940-925A-122	Sequence 122, App
	37	26	100.0	386	9	US-09-941-193A-122	Sequence 122, App
	38	26	100.0	7992	9	US-10-005-469-1	Sequence 1, Appl
	39	26	100.0	7992	9	US-10-005-469-2	Sequence 2, Appl
	40	26	100.0	7992	9	US-10-005-469-4	Sequence 4, Appl
	41	26	100.0	7992	9	US-10-005-469-5	Sequence 5, Appl
	42	26	100.0	7992	9	US-10-005-469-6	Sequence 6, Appl
	43	26	100.0	7995	9	US-10-005-469-3	Sequence 3, Appl
	44	26	100.0	8638	12	US-10-029-907-6	Sequence 6, Appl
	45	26	100.0	8638	12	US-10-029-907-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-10-087-631B-6
Sequence 6, Application US/10087631B
Publication No. US20030034372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087, 631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 26
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: sequence ST280 HCV-specifl
US-10-087-631B-6
OTHER INFORMATION: primer sequence
Query Match 100.0%; Score 26; DB 9; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1 GCAGAAAGCGTCTAGCCATGCGCTTA 26
1 GCAGAAAGCGTCTAGCCATGCGCTTA 26
RESULT 2
US-10-029-907-17
Sequence 17, Application US/10029907
Patent No. US20020142350A1
GENERAL INFORMATION:
APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
FILE REFERENCE: 13/083
CURRENT APPLICATION NUMBER: US/10/029, 907
CURRENT FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: 60/257,857
PRIOR FILING DATE: 2000-12-22
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 30
TYPE: DNA
ORGANISM: HCV
US-10-029-907-17

Query Match
Best Local Similarity 100.0%; Score 26; DB 12; Length 30;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
DB 3 GCAGAAAGCGCTAGCCATGGCGTTA 28

RESULT 3

US-10-087-631b-10
Sequence 10, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:

APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999

CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 241
TYPE: DNA

ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: DNA sequence derived by
OTHER INFORMATION: amplification of HCV type 1 using primers ST280 and ST778
US-10-087-631b-10

Query Match
Best Local Similarity 100.0%; Score 26; DB 9; Length 241;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
DB 1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 4

US-10-087-631b-13
Sequence 13, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:

APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999

CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 13
LENGTH: 241
TYPE: DNA

ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Description of artificial sequence: Amplicon sequence derived fr
OTHER INFORMATION: QS HCV (HCV amplification control having binding sites for ST280,
OTHER INFORMATION: and ST753) using primers ST280 and ST778
US-10-087-631b-13

Query Match
Best Local Similarity 100.0%; Score 26; DB 9; Length 241;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
DB 1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 5

US-10-087-631b-15
Sequence 15, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:

APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999

CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 15
LENGTH: 242
TYPE: DNA

ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Description of artificial sequence: Amplicon derived from ICS
OTHER INFORMATION: HCV (HCV-specific amplification control) using ST280 and ST778
US-10-087-631b-15

Query Match
Best Local Similarity 100.0%; Score 26; DB 9; Length 242;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
DB 1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 6

US-09-825-574-26
Sequence 26, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:

APPLICANT: Lyamichev, Victor I.
Brow, Mary Ann D.
Neri, Bruce P.

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.

NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-825-574-26

Query Match
Best Local Similarity 100.0%; Score 26; DB 10; Length 244;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTACCGCATGGCGTTA 26
DB 1 GCAGAAAGCGCTACCGCATGGCGTTA 26

RESULT 7
US-09-825-574-27
Sequence 27, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamlichev, Victor I.
Brow, Mary Ann D.
Fors, Lance P.
Neri, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"

SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-09-825-574-27

Query Match
Best Local Similarity 100.0%; Score 26; DB 10; Length 244;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTACCGCATGGCGTTA 26
DB 1 GCAGAAAGCGCTACCGCATGGCGTTA 26

RESULT 8
US-09-825-574-28
Sequence 28, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamlichev, Victor I.
Brow, Mary Ann D.
Fors, Lance P.
Neri, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-09-825-574-28

Query Match
Best Local Similarity 100.0%; Score 26; DB 10; Length 244;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTACCGCATGGCGTTA 26
DB 1 GCAGAAAGCGCTACCGCATGGCGTTA 26

RESULT 9
US-09-940-925A-123

Sequence 123, Application US/09940925A
Publication No. US20030054338A1
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
OLIVE, DAVID M.
LYAMICHEV, VICTOR I.
TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS
NUMBER OF SEQUENCES: 165
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 123:
SEQUENCE CHARACTERISTICS:
LENGTH: 281 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 123:
US-09-940-925A-123

Query Match 100.0%; Score 26; DB 9; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGCTTA 26
DB 11 GCAGAAAGCGTCTAGCCATGCGCTTA 36

RESULT 10
US-09-940-925A-125
Sequence 125, Application US/09940925A
Publication No. US20030054338A1
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
OLIVE, DAVID M.
LYAMICHEV, VICTOR I.
TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS
NUMBER OF SEQUENCES: 165
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 128:
SEQUENCE CHARACTERISTICS:
LENGTH: 281 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 128:

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 125:
SEQUENCE CHARACTERISTICS:
LENGTH: 281 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 125:
US-09-940-925A-125

Query Match 100.0%; Score 26; DB 9; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGCTTA 26
DB 11 GCAGAAAGCGTCTAGCCATGCGCTTA 36

RESULT 11
US-09-940-925A-128/c
Sequence 128, Application US/09940925A
Publication No. US20030054338A1
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
OLIVE, DAVID M.
LYAMICHEV, VICTOR I.
TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS
NUMBER OF SEQUENCES: 165
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 128:
SEQUENCE CHARACTERISTICS:
LENGTH: 281 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 128:

US-09-940-925A-128

Query Match 100.0%; Score 26; DB 9; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
DB 271 GCAGAAAGCGTCTAGCCATGGCGTTA 246

RESULT 12
US-09-940-925A-129/C

; Sequence 129, Application US/09940925A
; Publication No. US20030054338A1

GENERAL INFORMATION:

APPLICANT: BROW, MARY ANN D.
LYAMICHEV, VICTOR I.

OLIVE, DAVID M.

TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS

NUMBER OF SEQUENCES: 165

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MEDLEN & CARROLL

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/940.925A

FILING DATE: 10-Jun-2002

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: CARROLL, PETER G.

REGISTRATION NUMBER: 32,837

REFERENCE/DOCKET NUMBER: FORS-01756

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 129:

SEQUENCE CHARACTERISTICS:

LENGTH: 281 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 129:

US-09-940-925A-129

Query Match 100.0%; Score 26; DB 9; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
DB 271 GCAGAAAGCGTCTAGCCATGGCGTTA 246

RESULT 13
US-09-940-925A-131/C

; Sequence 131, Application US/09940925A
; Publication No. US20030054338A1

GENERAL INFORMATION:

APPLICANT: BROW, MARY ANN D.
LYAMICHEV, VICTOR I.

OLIVE, DAVID M.

TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF

PATHOGENS

NUMBER OF SEQUENCES: 165

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MEDLEN & CARROLL

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/940.925A

FILING DATE: 10-Jun-2002

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: CARROLL, PETER G.

REGISTRATION NUMBER: 32,837

REFERENCE/DOCKET NUMBER: FORS-01756

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 131:

SEQUENCE CHARACTERISTICS:

LENGTH: 281 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 131:

US-09-940-925A-131

Query Match 100.0%; Score 26; DB 9; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
DB 271 GCAGAAAGCGTCTAGCCATGGCGTTA 246

RESULT 14
US-09-941-193A-123

; Sequence 123, Application US/09941193A
; Publication No. US20030108873A1

GENERAL INFORMATION:

APPLICANT: BROW, MARY ANN D.
LYAMICHEV, VICTOR I.

OLIVE, DAVID M.

TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF

PATHOGENS

NUMBER OF SEQUENCES: 165

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MEDLEN & CARROLL

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/941.193A

FILING DATE: 28-Aug-2001

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: CARROLL, PETER G.

Db 11 GCAGAAAGCGCTAGCCATGCGCTTA 36
 Search completed: July 11, 2003, 15:01:56
 Job time: 71.164 secs

REGISTRATION NUMBER: 32,837
 REFERENCE/DOCKET NUMBER: FORS-01756
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 123:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 281 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 123:
 US-09-941-193A-123

Query Match 100.0%; Score 26; DB 9; Length 281;
 Best Local Similarity 100.0%; Pred. No. 0.0027;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGCGCTTA 26
 Db 11 GCAGAAAGCGCTAGCCATGCGCTTA 36

RESULT 15
 US-09-941-193A-125
 Sequence 125: Application US/09941193A
 Publication No. US20030108873A1

GENERAL INFORMATION:
 APPLICANT: BROW, MARY ANN D.
 LYAMICHEV, VICTOR I.
 OLIVE, DAVID M.

TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
 PATHOGENS

NUMBER OF SEQUENCES: 165
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MEDLEN & CARROLL
 STREET: 220 MONTGOMERY STREET, SUITE 2200
 CITY: SAN FRANCISCO
 STATE: CALIFORNIA
 COUNTRY: UNITED STATES OF AMERICA
 ZIP: 94104

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/941,193A
 FILING DATE: 28-Aug-2001
 CLASSIFICATION: <unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: CARROLL, PETER G.
 REGISTRATION NUMBER: 32,837
 REFERENCE/DOCKET NUMBER: FORS-01756
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 125:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 281 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 125:
 US-09-941-193A-125

Query Match 100.0%; Score 26; DB 9; Length 281;
 Best Local Similarity 100.0%; Pred. No. 0.0027;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGCGCTTA 26

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:13:45 ; Search time 467.474 Seconds
(without alignments)
900.761 Million cell updates/sec

Title: US-10-087-631b-6

Perfect score: 26

Sequence: 1 gccagaagcgtctagcgcgtcta 26

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estb1:
2: em_estb2:
3: em_estb3:
4: em_estb4:
5: em_estb5:
6: em_estb6:
7: em_estb7:
8: em_estb8:
9: em_estb9:
10: em_estb10:
11: em_estb11:
12: em_estb12:
13: em_estb13:
14: em_estb14:
15: em_estb15:
16: em_estb16:
17: em_estb17:
18: em_estb18:
19: em_estb19:
20: em_estb20:
21: em_estb21:
22: em_estb22:
23: em_estb23:
24: em_estb24:
25: em_estb25:
26: em_estb26:
27: em_estb27:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	20.2	77.7	549 12	BG443799 GA_Ea002
2	20.2	77.7	603 10	BE055455 GA_Ea000
3	20.2	77.7	610 13	BM358998 GA_Ea001
4	20.2	77.7	634 14	BQ404294 GA_Ed006
5	20.2	77.7	638 14	BQ401354 GA_Ed000
6	20.2	77.7	642 12	BG441998 GA_Ea001

7	20.2	77.7	694 12	BG440823 GA_Ea001
8	20.2	77.7	779 10	BE055455 GA_Ea000
9	20.2	77.7	840 10	BE278075 GA_Ed003
10	18.6	71.5	609 14	BQ839191 WHEA163.D
11	18.6	71.5	688 14	BQ405841 GA_Ed008
12	18.4	70.8	313 10	AM896937 RC4-NN005
13	18.2	70.0	240 9	AJ461009 AJ461009
14	18.2	70.0	752 12	A2559641 RPT-23-2
15	18	69.2	377 12	BF912055 IL2-UTR007
16	18	69.2	383 12	BF911367 IL2-UTR007
17	18	69.2	396 12	BF911282 IL2-UTR007
18	18	69.2	402 12	BF911293 IL2-UTR007
19	18	69.2	415 12	BF911279 IL2-UTR007
20	18	69.2	450 10	BE475077 sp72a05.Y
21	18	69.2	564 13	B1470037 sp74a05.Y
22	18	69.2	590 14	BQ742735 sag55009.
23	18	69.2	668 12	BG398045 AGENCOURT
24	18	69.2	883 14	BQ686524 AGENCOURT
25	17.8	68.5	326 14	BQ507679 EST615094
26	17.8	68.5	393 9	AU068851 AU068851
27	17.8	68.5	403 9	AU076005 AU076005
28	17.8	68.5	421 9	AU162743 AU162743
29	17.8	68.5	435 14	D49161
30	17.8	68.5	595 17	A2865010 2M0174K22
31	17.6	67.7	344 9	A1255537 u155g04.Y
32	17.6	67.7	520 13	B1188143 B1188143
33	17.6	67.7	520 13	BM168779 EST571302
34	17.6	67.7	549 10	BE632435 uv56c06.Y
35	17.6	67.7	563 12	BG146531 mad93b06.
36	17.6	67.7	564 13	B1184088 B1184088
37	17.6	67.7	566 13	BM660882 952042F08
38	17.6	67.7	637 10	BB625273 BB625273
39	17.6	67.7	662 17	A2612987 1M0441P12
40	17.6	67.7	695 17	BH558329 BOGHAN27R
41	17.6	67.7	726 14	BM937432 UT-M-CD1
42	17.6	67.7	735 17	BH520932 BOGGA53TF
43	17.6	67.7	807 17	BH597113 BOGDY77TR
44	17.6	67.7	812 13	B1218614 602937835
45	17.6	67.7	833 17	BH490369 BOHHE31TF

ALIGNMENTS

RESULT 1
LOCUS BG443799 549 bp mRNA linear EST 15-MAR-2001
DEFINITION GA_Ea0022C04f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Ea0022C04f, mRNA sequence.

ACCESSION BG443799.1 GI:13353451

KEYWORDS EST.

ORGANISM Gossypium arboreum.

REFERENCE Gossypium arboreum.

AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

TITLE Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Unpublished (2000)

CONTACT: Wing RA

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: twing@clemson.edu

Seq primer: TTAATGACATCACTATAGCG

High quality sequence stop: 542.

Location/Qualifiers

source

1. 549

/organism="Gossypium arboreum"

/strain="AKA"

/db_xref="taxon:29729"

/clone_lib="Gossypium arboreum 7-10 dpa fiber library"

/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"

/lab_host="E. coli"

/note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT

156 a 119 c 125 g 148 t 1 others

Query Match

Best Local Similarity 77.7%; Score 20.2; DB 12; Length 549;
Matches 22; Conservativity 0; Mismatches 3; Indels 0; Gaps 0;

Db

1 GCAGAAAGCCTAGCCATGCGCTT 25
6 GCAGAAAGCCTAGCCATGCGCTT 30

RESULT 2

BE055455

LOCUS

603 bp mRNA linear EST 07-MAR-2001

ACCESSION

BE055455

VERSION

BE055455.2 GI:13244672

KEYWORDS

EST

SOURCE

Gossypium arboreum

ORGANISM

Gossypium arboreum

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Malvales; Malvaceae; Gossypium.

AUTHORS

Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry

TITLE

An integrated analysis of the genetics, development, and evolution

JOURNAL

Unpublished (2000)

COMMENT

On Jun 8, 2000 this sequence version replaced gi:8382512.

CONTACT

Wing, R.A.

Clemson University

Clemson University

100 Jordan Hall,

Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: TAATACGACTCACTATAGCG

High quality sequence stop: 600.

Location/Qualifiers

1. 603

/organism="Gossypium arboreum"

/strain="AKA"

/db_xref="taxon:29729"

/clone_lib="Gossypium arboreum 7-10 dpa fiber library"

/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"

/lab_host="E. coli"

/note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI"

172 a 121 c 132 g 163 t 15 others

BASE COUNT

172 a 121 c 132 g 163 t 15 others

Query Match

Best Local Similarity 77.7%; Score 20.2; DB 10; Length 603;
Matches 22; Conservativity 0; Mismatches 3; Indels 0; Gaps 0;

Db

1 GCAGAAAGCCTAGCCATGCGCTT 25

TITLE

An integrated analysis of the genetics, development, and evolution

Db

28 GCAGAAAGCCTAGCCATGCGCTT 52

RESULT 3

BM358998

LOCUS

610 bp mRNA linear EST 09-JAN-2002

DEFINITION

Gossypium arboreum 7-10 dpa fiber library Gossypium

ACCESSION

BM358998

VERSION

BM358998.1 GI:18099744

KEYWORDS

EST

SOURCE

Gossypium arboreum

ORGANISM

Gossypium arboreum

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Malvales; Malvaceae; Gossypium.

AUTHORS

Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry

TITLE

An integrated analysis of the genetics, development, and evolution

JOURNAL

Unpublished (2000)

COMMENT

On Jun 8, 2000 this sequence version replaced gi:8382512.

CONTACT

Wing, R.A.

Clemson University

Clemson University

100 Jordan Hall,

Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: TAATACGACTCACTATAGCG

High quality sequence stop: 407.

Location/Qualifiers

1. 610

/organism="Gossypium arboreum"

/strain="AKA"

/db_xref="taxon:29729"

/clone_lib="Gossypium arboreum 7-10 dpa fiber library"

/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"

/lab_host="E. coli"

/note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI"

170 a 130 c 139 g 171 t

BASE COUNT

170 a 130 c 139 g 171 t

Query Match

Best Local Similarity 77.7%; Score 20.2; DB 13; Length 610;
Matches 22; Conservativity 0; Mismatches 3; Indels 0; Gaps 0;

Db

1 GCAGAAAGCCTAGCCATGCGCTT 25

TITLE

An integrated analysis of the genetics, development, and evolution

JOURNAL

Unpublished (2000)

COMMENT

On Jun 8, 2000 this sequence version replaced gi:8382512.

CONTACT

Wing, R.A.

Clemson University

Clemson University

100 Jordan Hall,

Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: TAATACGACTCACTATAGCG

High quality sequence stop: 407.

Location/Qualifiers

1. 610

/organism="Gossypium arboreum"

/strain="AKA"

/db_xref="taxon:29729"

/clone_lib="Gossypium arboreum 7-10 dpa fiber library Gossypium

/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"

/lab_host="E. coli"

/note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI"

172 a 121 c 132 g 163 t 15 others

BASE COUNT

172 a 121 c 132 g 163 t 15 others

JOURNAL
COMMENT

of the cotton fiber
unpublished (2000)
Contact: Ming RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total High Quality bases = 593
Seq primer: TAAATGACCTCCTATAGG
High quality sequence start: 6
High quality sequence stop: 632.
Location/Qualifiers
1. 634

FEATURES
SOURCE

/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ed0015G07f"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT

188 a 124 c 140 g 176 t 6 others

ORIGIN

Query Match 77.7%; Score 20.2; DB 14; Length 634;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTACCATGCGCTT 25
|||||
DB 28 GCAGAAAGCGTCTACCATGACTTT 52

RESULT 5
LOCUS

DEFINITION BQ401354 638 bp mRNA linear EST 22-MAY-2002
GA_Ed0015G07f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Ed0015G07f, mRNA sequence.

ACCESSION BQ401354
VERSION BQ401354.1 GI:21089041
KEYWORDS EST
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum.

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 638)

Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
D., Wood, T.C., Leslie, A. and Wilkins, T.A.

An integrated analysis of the genetics, development, and evolution
of the cotton fiber

unpublished (2000)

JOURNAL
COMMENT

Contact: Ming RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total High Quality bases = 584
Seq primer: TAAATGACCTCCTATAGG
High quality sequence stop: 635.
Location/Qualifiers
1. 638

FEATURES
SOURCE

/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ed0015G07f"

/clone="lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT

188 a 125 c 144 g 181 t

ORIGIN

Query Match 77.7%; Score 20.2; DB 14; Length 638;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTACCATGCGCTT 25
|||||
DB 22 GCAGAAAGCGTCTACCATGACTTT 46

RESULT 6
LOCUS

DEFINITION BQ441998 642 bp mRNA linear EST 15-MAR-2001
GA_Ea0015G07f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Ea0015G07f, mRNA sequence.

ACCESSION BQ441998
VERSION BQ441998.1 GI:13351650
KEYWORDS EST
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 642)

Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
D., Wood, T.C., Leslie, A. and Wilkins, T.A.

An integrated analysis of the genetics, development, and evolution
of the cotton fiber

unpublished (2000)

JOURNAL
COMMENT

Contact: Ming RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAAATGACCTCCTATAGG
High quality sequence stop: 642.
Location/Qualifiers
1. 642

FEATURES
SOURCE

/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0015G07f"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT

187 a 128 c 146 g 181 t

ORIGIN

Query Match 77.7%; Score 20.2; DB 12; Length 642;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTACCATGCGCTT 25
|||||
DB 31 GCAGAAAGCGTCTACCATGACTTT 55

RESULT 7
LOCUS

DEFINITION BQ440823 694 bp mRNA linear EST 15-MAR-2001
GA_Ea0015G07f Gossypium arboreum 7-10 dpa fiber library Gossypium

ACCESSION arboresum cDNA clone GA_Ea0010F07f, mRNA sequence.
 VERSION BG440823.1 GI:13350475
 KEYWORDS EST
 SOURCE Gossypium arboreum.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II; Malvales; Malvaceae; Gossypium.
 REFERENCE 1 (bases 1 to 694)
 AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber
 JOURNAL Unpublished (2000)
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: TAATACGACTCATGATGAGG
 High quality sequence stop: 353.
 FEATURES
 SOURCE location/Qualifiers
 1..694
 /organism="Gossypium arboreum"
 /strain="AKA"
 /cultivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_Ea0010F07f"
 /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
 /tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
 /lab_host="E. coli"
 /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 191 a 146 c 154 g 203 t
 ORIGIN
 Query Match 77.7%; Score 20.2; DB 12; Length 694;
 Best Local Similarity 88.0%; Fred. No. 1.1e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCCTGATGCCATGCGTT 25
 ||||||| ||||||| |||||
 18 GCAGAAAGCCTGATGCCATGACTT 42
 RESULT 8 779 bp mRNA linear EST 07-MAR-2001
 BE055545
 LOCUS GAL_Ea0004L16f Gossypium arboreum 7-10 dpa fiber library Gossypium
 DEFINITION arboresum cDNA clone GA_Ea0004L16f, mRNA sequence.
 ACCESSION BE055545
 VERSION BE055545.2 GI:13244556
 KEYWORDS EST
 SOURCE Gossypium arboreum.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II; Malvales; Malvaceae; Gossypium.
 REFERENCE 1 (bases 1 to 779)
 AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber
 JOURNAL Unpublished (2000)
 COMMENT On Jun 8, 2000 this sequence version replaced gi:8382602.
 Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288

Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: TAATACGACTCATGATGAGG
 High quality sequence stop: 613.
 FEATURES
 SOURCE location/Qualifiers
 1..779
 /organism="Gossypium arboreum"
 /strain="AKA"
 /cultivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_Ea0004L16f"
 /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
 /tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
 /lab_host="E. coli"
 /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 220 a 158 c 177 g 217 t
 ORIGIN
 Query Match 77.7%; Score 20.2; DB 10; Length 779;
 Best Local Similarity 88.0%; Fred. No. 1.1e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCCTGATGCCATGCGTT 25
 ||||||| ||||||| |||||
 27 GCAGAAAGCCTGATGCCATGACTT 51
 RESULT 9 840 bp mRNA linear EST 07-MAR-2001
 BE278075
 LOCUS GA_EB0038A09f Gossypium arboreum 7-10 dpa fiber library Gossypium
 DEFINITION arboresum cDNA clone GA_EB0038A09f, mRNA sequence.
 ACCESSION BE278075
 VERSION BE278075.1 GI:11209061
 KEYWORDS EST
 SOURCE Gossypium arboreum.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II; Malvales; Malvaceae; Gossypium.
 REFERENCE 1 (bases 1 to 840)
 AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber
 JOURNAL Unpublished (2000)
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: TAATACGACTCATGATGAGG
 High quality sequence start: 4
 High quality sequence stop: 535.
 FEATURES
 SOURCE location/Qualifiers
 1..840
 /organism="Gossypium arboreum"
 /strain="AKA"
 /cultivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_EB0038A09f"
 /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
 /tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
 /lab_host="E. coli"
 /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 198 a 194 c 148 g 298 t
 ORIGIN
 Query Match 77.7%; Score 20.2; DB 12; Length 840;

Best Local Similarity 88.0%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTT 25
Db 30 GCAGAAAGCGTCTAGCCATGGCGTT 54

RESULT 10
B0839191

LOCUS B0839191 609 bp mRNA linear EST 08-AUG-2002
DEFINITION WHE4163_D05.G0925 wheat CS whole plant cDNA library Triticum aestivum cDNA clone WHE4163_D05.G09, mRNA sequence.

ACCESSION B0839191

VERSION B0839191.1 GI:22143513

KEYWORDS EST

SOURCE bread wheat.

ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae

REFERENCE 1 (bases 1 to 609)

AUTHORS Anderson O.D., Akhunov, E., Chao, S., Crossman, C., Deal, K., Dvorak, J., Lazo, G.R., Rausch, C.J., Wilson, C., and Woo, J.

TITLE The structure and function of the expressed portion of the wheat genomes - Chinese Spring whole plant cDNA library

JOURNAL Unpublished (2002)

COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818

FEATURES

source Email: andersn@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: SK primer.

FEATURES

source location/Qualifiers

1..609
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE4163_D05.G09"
/clone_id="wheat CS whole plant cDNA library"
/tissue_type="Roots, leaves, crown, stem and sheath"
/dev_stage="Adult"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid
bluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Plant
tissues from wheat cv. CS grown to full tillering stage in
greenhouse were collected at University of California,
Davis (Jan Dvorak's lab, K. Deal and E. Akhunov). Total
RNA was prepared from leaves (young leaf and third leaf),
whole roots, crown, stem and sheath tissues, and then
equal quantities of RNA were pooled from the these
samples. PolyA was purified from the pooled RNA, a cDNA
library was made, and the cDNA clones were in vivo excised
to give phagemid SK(-) phagemids in J. Dvorak's lab (E.
Akhunov, J. Dvorak) at the University of California,
Davis. Colony plating, plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."

BASE COUNT 109 a 183 c 231 g 86 t
ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 609;
Best Local Similarity 84.0%; Pred. No. 5.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTT 25
Db 435 GCAGAAAGCGTCTAGCCATGGCGTT 459

RESULT 11
B0405841

LOCUS B0405841 688 bp mRNA linear EST 22-MAY-2002
DEFINITION GA_Ed0087E10f Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum cDNA clone GA_Ed0087E10f, mRNA sequence.

ACCESSION B0405841

VERSION B0405841.1 GI:21093528

KEYWORDS EST

SOURCE Gossypium arboreum.

ORGANISM Gossypium arboreum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium

REFERENCE 1 (bases 1 to 688)

AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.

TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber

JOURNAL Unpublished (2000)

COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293

FEATURES

source Email: rwing@clemson.edu
Total High Quality bases = 442
Seq primer: TAATGACCTCCTATGAGG
High quality sequence start: 33
High quality sequence stop: 661.
location/Qualifiers

1..688
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ed0087E10f"
/clone_id="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="Vector: pRK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 201 a 134 c 149 g 204 t
ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 688;
Best Local Similarity 84.0%; Pred. No. 5.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTT 25
Db 12 GCAGAAAGCGTCTAGCCATGGCGTT 36

RESULT 12
AM896937 313 bp mRNA linear EST 24-MAY-2000
LOCUS AM896937
DEFINITION RC4-NN0055-060400-011-f09 NN0055 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM896937
VERSION AM896937.1 GI:8061142
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M.R., Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

REFERENCE
AUTHORS

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=6t2-RCA-NN0055-060400-011-t096t3-2000-04-06&4-1)
 Seq primer: puc 18 forward
 High quality sequence, start: 21
 High quality sequence, stop: 313.
FEATURES
SOURCE
 1. 313
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN0055"
 /dev_stage="Adult"
 /note="Organ: nervous normal; Vector: puc18; Site: 1; SmaI; Site: 2; SmaI; A mini-library was made by cloning products derived from ORNSTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 64 a 85 c 83 g 81 t

BASE COUNT

ORIGIN
 Query Match 70.8%; Score 18.4; DB 10; Length 313;
 Best Local Similarity 95.0%; Pred. No. 4.5e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 AGCGTCAGCCATGCGCTTA 26
 ||||||||||||||||||||
 68 AGCGTCAGCCATGCGCTGA 87

RESULT 13
LOCUS AJ461009 240 bp mRNA linear EST 24-MAY-2002
DEFINITION AJ461009 S00002 Hordeum vulgare cDNA clone S0000200049A08F1, mRNA
ACCESSION AJ461009
VERSION AJ461009
KEYWORDS EST
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticeae; Hordeum.
 1 (bases 1 to 240)
 Saren, A.-M., Tanskanen, J., Paulin, L. and Schulman, A.H.
 Barley EST 5
 Unpublished (2002)
 Contact: Schulman AH
 Institute of Biotechnology
 University of Helsinki
 P.O. Box 56 (Vilkkinkaari 6A), University of Helsinki FIN-00014,
 Finland.
FEATURES
SOURCE
 1. 240
 Location/Qualifiers
 /organism="Hordeum vulgare"
 /cultivar="Saana"
 /db_xref="taxon:4513"
 /clone="S0000200049A08F1"
 /clone_lib="S00002"

BASE COUNT /dev stage="Embryo"
 /note="1 day after pollination"
ORIGIN 63 a 59 c 57 g 61 t

QY 1 GCAGAAAGCGCTAGCCATGCGC 23
 ||||||||||||||||||||
 45 GAAGAAAGCGCTAGCCATGCGC 67

RESULT 14
LOCUS A2559641 752 bp DNA linear GSS 20-NOV-2000
DEFINITION RPCI-23-205015.TV RPCI-23 Mus musculus genomic clone RPCI-23-205015
 / DNA sequence.
ACCESSION A2559641
VERSION A2559641
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 752)
 Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akiret
 'B., Levins, M., McGann, S., Ysegaire, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other GSSs: RPCI-23-205015.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buitalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buitalo.edu/orderingframe.htm)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
 http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html
 Plate: 205 row: 0 column: 15
 Seq primer: T7
 Class: BAC ends.
FEATURES
SOURCE
 1. 752
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-205015"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1;
 EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies).
BASE COUNT 216 a 175 c 159 g 202 t

QY 1 GCAGAAAGCGCTAGCCATGCGC 23
 ||||||||||||||||||||
 45 GAAGAAAGCGCTAGCCATGCGC 67

RESULT 14
LOCUS A2559641 752 bp DNA linear GSS 20-NOV-2000
DEFINITION RPCI-23-205015.TV RPCI-23 Mus musculus genomic clone RPCI-23-205015
 / DNA sequence.
ACCESSION A2559641
VERSION A2559641
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 752)
 Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akiret
 'B., Levins, M., McGann, S., Ysegaire, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other GSSs: RPCI-23-205015.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buitalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buitalo.edu/orderingframe.htm)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
 http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html
 Plate: 205 row: 0 column: 15
 Seq primer: T7
 Class: BAC ends.
FEATURES
SOURCE
 1. 752
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-205015"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1;
 EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies).
BASE COUNT 216 a 175 c 159 g 202 t

QY 1 GCAGAAAGCGCTAGCCATGCGC 23
 ||||||||||||||||||||
 45 GAAGAAAGCGCTAGCCATGCGC 67

RESULT 14
LOCUS A2559641 752 bp DNA linear GSS 20-NOV-2000
DEFINITION RPCI-23-205015.TV RPCI-23 Mus musculus genomic clone RPCI-23-205015
 / DNA sequence.
ACCESSION A2559641
VERSION A2559641
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 752)
 Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akiret
 'B., Levins, M., McGann, S., Ysegaire, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other GSSs: RPCI-23-205015.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buitalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buitalo.edu/orderingframe.htm)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
 http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html
 Plate: 205 row: 0 column: 15
 Seq primer: T7
 Class: BAC ends.
FEATURES
SOURCE
 1. 752
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-205015"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1;
 EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies).
BASE COUNT 216 a 175 c 159 g 202 t

Db 701 GCAGAAAGTGTCTAGTCAGGCG 723

RESULT 15

BF912055/c 377 bp mRNA linear EST 18-JAN-2001
 LOCUS IL2-UT0073-121100-232-D04 UT0073 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF912055
 ACCESSION BF912055
 VERSION BF912055.1 GI:12303513
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL2&tl2=IL2-UT0073-121100-232-D04&tl3=2000-11-12&tl4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 349.
 Location/Qualifiers

FEATURES
 Source 1..377
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="UT0073"
 /dev_stage="Adult"
 /note="Organ: uterus; tumor; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 98 a 66 c 76 g 136 t 1 others
 ORIGIN

Query Match 69.2%; Score 18; DB 12; Length 377;
 Best Local Similarity 80.8%; Pred. No. 7.7e+02;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GCAGAAAGCTGTAGCCATGCGCTTA 26
 ||||| ||||| ||||| |||||
 Db 322 GCAGAAAGCTCAACCCATGTAGTTA 297

Search completed: July 11, 2003, 02:25:50
 Job time : 474.474 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 59.0831 Seconds

(without alignments)
1067.241 Million cell updates/sec

Title: US-10-087-631B-7

Perfect score: 28

Sequence: 1 gcaagaccctatcgagcagatcacca 28

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues 4370478

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries.

Database : N.Geneseq_101002.*

1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*

2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*

3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*

4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*

5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*

6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*

7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*

8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*

9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*

10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*

11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*

12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*

13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*

14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*

15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*

16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*

17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*

18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*

19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*

20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*

21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*

22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*

23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*

24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	100.0	28	18	AAAT67194
2	28	100.0	28	19	AAV59059
3	28	100.0	28	22	AAH25414
4	28	100.0	57	14	AAO63223
5	28	100.0	186	17	AAAT11272
6	28	100.0	209	24	ABLA1919
7	28	100.0	232	19	AAV70460
8	28	100.0	232	24	ABLA6070
9	28	100.0	239	19	AAV70455

C 10	28	100.0	239	19	AAV70459	Partial sequence o
C 11	28	100.0	239	24	ABLA6065	Hepatitis C virus
C 12	28	100.0	239	24	ABLA6069	Hepatitis C virus
C 13	28	100.0	240	19	AAV70457	Partial sequence o
C 14	28	100.0	240	19	AAV70456	Partial sequence o
C 15	28	100.0	240	19	AAV70458	Partial sequence o
C 16	28	100.0	240	24	ABLA6066	Hepatitis C virus
C 17	28	100.0	240	24	ABLA6067	Hepatitis C virus
C 18	28	100.0	240	24	ABLA6068	Hepatitis C virus
C 19	28	100.0	244	19	AAV70454	Partial sequence o
C 20	28	100.0	244	19	AAV70449	HCV subtype 1a PCR
C 21	28	100.0	244	19	AAV70450	HCV subtype 1b PCR
C 22	28	100.0	244	19	AAV70451	HCV subtype 2c PCR
C 23	28	100.0	244	19	AAV70452	HCV subtype 3a PCR
C 24	28	100.0	244	24	ABLA6059	Hepatitis C virus
C 25	28	100.0	244	24	ABLA6060	Hepatitis C virus
C 26	28	100.0	244	24	ABLA6061	Hepatitis C virus
C 27	28	100.0	244	24	ABLA6062	Hepatitis C virus
C 28	28	100.0	244	24	ABLA6064	Hepatitis C virus
C 29	28	100.0	252	13	AAO31071	HCV-1 genotype GII
C 30	28	100.0	252	13	AAO31072	HCV-1 genotype GII
C 31	28	100.0	252	13	AAO31078	HCV-1 genotype GII
C 32	28	100.0	252	13	AAO31079	HCV-1 genotype GII
C 33	28	100.0	252	13	AAO31067	HCV-1 genotype GI
C 34	28	100.0	252	13	AAO31067	HCV-1 genotype GI
C 35	28	100.0	252	13	AAO31068	HCV-1 genotype GI
C 36	28	100.0	252	13	AAO31069	HCV-1 genotype GI
C 37	28	100.0	252	13	AAO31070	HCV-1 genotype GI
C 38	28	100.0	252	13	AAO31080	HCV-1 genotype GIV
C 39	28	100.0	252	13	AAO31081	HCV-1 genotype GIV
C 40	28	100.0	256	13	AAO32981	HCV EI 5' non-codi
C 41	28	100.0	256	20	AAH84004	Hepatitis C virus
C 42	28	100.0	267	20	AAH16761	Hepatitis C virus
C 43	28	100.0	267	24	AAH79975	Hepatitis C virus
C 44	28	100.0	278	24	AAH79973	Hepatitis C virus
C 45	28	100.0	279	24	ABA01119	Hepatitis C virus

ALIGNMENTS

RESULT 1	
AAAT67194	
ID	AAAT67194 standard; DNA; 28 Bp.
AC	AAAT67194;
AC	13-FEB-1998 (first entry)
DT	Hepatitis C virus (HCV) RNA amplification primer ST778AA.
XX	Hepatitis C virus; HCV; ST778AA; reverse transcription PCR; RT-PCR;
XX	detection; PCR primer; ss.
KW	Synthetic.
OS	
XX	
PN	EP776981-A2.
XX	04-JUN-1997.
PD	
XX	
XX	21-NOV-1996; 96EP-0118704.
PF	
XX	29-NOV-1995; 95US-0007739.
PR	
XX	(HOFF) HOFFMANN LA ROCHE & CO AG F.
PA	Tsang SY;
XX	
PI	WPI; 1997-291296/27.
XX	
DR	Oligonucleotide primers for hepatitis C virus RNA amplification
XX	by polymerase chain reaction
PT	
XX	

PS Claim 1, Page 12, 16pp; English.

CC This downstream primer ST778AA is used in the amplification of the

CC Hepatitis C virus (HCV) RNA by reverse transcription PCR. This is used

CC to amplify a 250 base pair product from the 5' untranslated region of

CC the HCV genome. This can be used to detect HCV in a sample with increased

CC sensitivity. Amplification of HCV nucleic acid using this primer is up to

CC 100 times more efficient than amplification with prior art primers.

XX

SQ Sequence 28 BP; 10 A; 10 C; 5 G; 3 T; 0 other;

Query Match 100.0%; Score 28; DB 18; Length 28;

Best Local Similarity 100.0%; Pred. No. 0.0028;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCTATCAGCAGTACCA 28

DB 1 GCAAGCACCTATCAGCAGTACCA 28

RESULT 2

AAV59059 standard; DNA; 28 BP.

XX AAV59059;

AC AAV59059;

XX

DT 07-JAN-1999 (first entry)

XX

DE Primer ST778AA for HCV fragment.

XX

KW PCR primer; HCV; nucleic acid amplification; ss.

XX

OS Synthetic.

OS Human cytomegalovirus.

XX

FT Key Location/Qualifiers

FT modified_base 25

FT /tag= a

FT /note= "optionally benzylated"

FT modified_base 27

FT /tag= b

FT /note= "optionally benzylated"

FT modified_base 28

FT /tag= c

FT /note= "optionally benzylated; methylated, or

FT nitrobenzylated"

XX

PN EP866071-A2.

XX

PD 23-SEP-1998.

XX

PF 12-MAR-1998; 98EP-0104461

XX

PR 20-MAR-1997; 97US-0041127

XX

PA (HOFF) HOFFMANN LA ROCHE & CO AG F.

XX

PI Will SG, Young KKY;

XX

DR WPI; 1998-482929/42.

XX

PT Oligo-nucleotide(s) containing N-substituted nucleotide useful as

PT primers for nucleic acid amplification

XX

PS Example 6; Page 16; 38pp; English.

XX

CC This sequence represents a primer for a fragment of HCV, and is an

CC example of an oligonucleotide of the invention. The oligonucleotides of

CC the invention are of the formula 5'-SI-Nu-3' or 5'-SI-Nu-S2-3', where

CC SI is a sequence of 5-50 nucleotides; S2 is a sequence of 1-3

CC nucleotides; and Nu is a nucleotide with a purine or pyrimidine base

CC having an exocyclic amino group substituted by CHR1R2; R1, R2 are H,

CC 1-10C alkyl, alkoxy, optionally substituted phenyl, phenoxy or optionally

CC substituted naphthyl. The oligonucleotides are useful as primers for

CC nucleic acid amplification, preferably by polymerase chain reaction. Use

CC of the modified primers reduces non-specific amplification, especially

CC primer dimer formation, with a concomitant increase in the yield of the

CC intended target.

XX

SQ Sequence 28 BP; 10 A; 10 C; 5 G; 3 T; 0 other;

Query Match 100.0%; Score 28; DB 19; Length 28;

Best Local Similarity 100.0%; Pred. No. 0.0028;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCTATCAGCAGTACCA 28

DB 1 GCAAGCACCTATCAGCAGTACCA 28

RESULT 3

AAH25414 standard; DNA; 28 BP.

XX AAH25414;

AC AAH25414;

XX

DT 22-AUG-2001 (first entry)

XX

DE Reverse PCR primer used to amplify a HCV DNA fragment.

XX

KW Magnetic glass particle; nucleic acid purification; PCR primer; ss.

XX

OS Hepatitis C virus.

OS

FT Key Location/Qualifiers

FT modified_base 28

FT /tag= a

FT /note= "derivatisation with a p-(t-butyl)benzyl residue"

XX

PN WO200137291-A1.

XX

PD 25-MAY-2001.

XX

PF 17-NOV-2000; 2000WO-EP11459.

XX

PR 17-NOV-1999; 99EP-0122853.

XX

PR 12-MAY-2000; 2000EP-0110165.

XX

PA (HOFF) ROCHE DIAGNOSTICS GMBH.

XX

PI Weindel K, Riedling M, Geiger A;

XX

DR WPI; 2001-381247/40.

XX

PT Novel composition of magnetic glass particles for purification of DNA

PT or RNA in automated processes

XX

PS Example 7; Page 98; 105pp; English.

XX

CC The specification describes a composition of magnetic glass particles,

CC which contain at least one magnetic object with a mean diameter between

CC 5-500 nm. The composition is useful for the purification of nucleic

CC acids. The composition can be used to process large quantities of

CC nucleic acid samples, because it does not involve the particles being

CC centrifuged or the fluids being drawn through glass fiber filters.

CC PCR primers AAH25413-14 were used to amplify HCV DNA fragments. The

CC amplified fragment can be purified using the method of the invention.

XX

SQ Sequence 28 BP; 10 A; 10 C; 5 G; 3 T; 0 other;

Query Match 100.0%; Score 28; DB 22; Length 28;

Best Local Similarity 100.0%; Pred. No. 0.0028;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCTATCAGCAGTACCA 28

DB 1 GCAAGCACCTATCAGCAGTACCA 28

```

Db          1  GCAGGACACCCCTATCAGGCAGTACCAACAA 28

RESULT 4
AA063223/c
ID  AA063223 standard; RNA; 57 BP.
XX
XX
AC  AA063223;
XX
XX  13-JUN-1994 (first entry)
DT
XX  Hepatitis C virus probe target region.
DE
XX  Detection; HCV; 11:2 probe design.
XX
XX  Hepatitis C virus.
OS
XX  W093324656-A.
PN
XX  09-DEC-1993.
PD
XX
XX  24-MAY-1993; 93WO-US04863.
PE
XX
XX  29-MAY-1992; 92US-0891543.
PR
XX
XX  (ABBO ) ABBOTT LAB.
PA
XX
XX  Carrino JJ, Marshall RL, Sustachek JC;
PI
XX
XX  WPI; 1993-405844/50.
DR
XX
XX  Amplifying known RNA target for use in diagnosis of HIV and HCV.
PT  Infection - by treating sample RNA with oligo-nucleotide probe,
PT  extending probe by reverse transcription of target, dissociating
PT  probe from target, hybridising 2nd probe with 1st, etc.
XX
XX  Example 8; Page 26; 49pp; English.
PS
XX
XX  The sequence is that of the target region of probes (AA053257-053260)
CC  used in the detection of hepatitis C virus (HCV) using a 11:2 probe
CC  design. It corresponds to positions 246-302 of the 5' UTR of the
CC  HPC6HDMK sequence.
XX
XX  Sequence 57 BP; 9 A; 9 C; 23 G; 16 T; 0 other;
SO

Query Match          100.0%; Score 28; DB 14; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.0031;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GCAGGACACCCCTATCAGGCAGTACCAACAA 28
        |||||||||||||||||||||||||||
DB      54  GCAAGCACCCCTATCAGGCAGTACCAACAA 27

RESULT 5
AA011272
ID  AA011272 standard; RNA; 186 BP.
XX
XX  AA011272;
AC
XX
XX  26-JUN-1996 (first entry)
DT
XX
XX  Hepatitis C virus partial 5'-UTR antisense RNA AS15.
DE
XX
XX  Antisense; therapy; complementary; HCV; 5'-untranslated region;
RW  hepatitis C virus; inhibition; infection; treatment; stem-loop;
XX  clone 2-1; ss.
XX
XX  Hepatitis C virus.
OS
XX
XX  JP07303485-A.
PN
XX
XX  21-NOV-1995.

```

XX	13-MAY-1994;	94JP-0124609.
PF		
XX	13-MAY-1994;	94JP-0124609.
PR		
XX	(TOFU) TONEN CORP.	
PA		
XX	WPI; 1996-035187/04.	
DR		
XX	Hepatitis C virus (HCV) anti:sense RNA - inhibits HCV structural	
PT	gene expression in vivo for treatment of HCV infection	
XX		
PS	Claim 2; Page 10; 12pp; Japanese.	
XX		
CC	The present sequence is a specifically claimed example of RNA that	
CC	is complementary (i.e. antisense) to part of the 5'-untranslated	
CC	region of the hepatitis C virus genome sequence contained in clone	
CC	2-1. The 5'-UTR includes several stem-loop sequences. The antisense	
CC	RNA is useful for inhibiting expression of HCV structural genes and	
CC	thereby inhibiting viral replication in vivo. The antisense therapy	
CC	can be used in addition to conventional interferon treatment of HCV	
CC	infections.	
XX		
SQ	Sequence 186 BP; 41 A; 65 C; 48 G; 32 U; 0 other;	
Query Match	100.0%; Score 28; DB 17; Length 186;	
Best Local Similarity	89.3%; Pred. No. 0.0037;	
Matches 25; Conservative 3; Mismatches 0; Indels 0; Gaps 0;		
OY	1 GCAGCACCCTATCAGCAGTACCAACA 28	
	: : : : : : : : :	
DB	10 GCAGCACCCTAUCAGCAGUACCAACA 37	
RESULT 6		
ABL41919/C		
ID	ABL41919; standard; RNA; 209 BP.	
XX		
AC	ABL41919;	
XX		
DT	11-JUN-2002 (first entry)	
XX		
DE	DNA comprising consensus endodogmatic sequence/exodogmatic sequence.	
XX		
KW	cellular organism; pathogen; retroviral particle; probe; ss.	
XX		
OS	Synthetic.	
XX		
PH		
FT	Key	Location/Qualifiers
FT	stem_loop	5..49
FT		/*tag= a
FT	misc_structure	5..7
FT		/*tag= b
FT		/note= "these bases bind to bases 47-49"
FT	misc_structure	10..14
FT		/*tag= c
FT		/note= "these bases bind to bases 39-43"
FT	misc_structure	39..43
FT		/*tag= d
FT		/note= "these bases bind to bases 10-14"
FT	stem_loop	92..104
FT		/*tag= e
FT	misc_structure	92..96
FT		/*tag= f
FT		/note= "these bases bind to bases 100-104"
FT	misc_structure	100..104
FT		/*tag= g
FT		/note= "these bases bind to bases 92-96"
FT	stem_loop	105..118
FT		/*tag= h
FT	misc_structure	105..108
FT		/*tag= i
FT		/note= "these bases bind to bases 115-118"
FT		

```

FT misc_structure 116..118
FT /tag= j
FT /note= "theses bases bind to bases 105-108"
FT stem_loop 153..208
FT /tag= k
FT misc_structure 153..157
FT /tag= l
FT /note= "theses bases bind to bases 204-208"
FT misc_structure 158..160
FT /tag= m
FT /note= "theses bases bind to bases 193-195"
FT misc_structure 162..167
FT /tag= n
FT /note= "theses bases bind to bases 186-191"
FT misc_structure 171..173
FT /tag= o
FT /note= "theses bases bind to bases 182-184"
FT misc_structure 182..184
FT /tag= p
FT /note= "theses bases bind to bases 171-173"
FT misc_structure 186..191
FT /tag= q
FT /note= "theses bases bind to bases 162-167"
FT misc_structure 193..195
FT /tag= r
FT /note= "theses bases bind to bases 158-160"
FT misc_structure 204..208
FT /tag= s
FT /note= "theses bases bind to bases 204-208"
FT WO200202803-A2.
FT 10-JAN-2002.
FT 18-JUN-2001; 2001WC-CH00381.
FT 03-JUL-2000; 2000CH-0001311.
FT (HAFN/) HAFNER W. A.
FT (MEMO/) MEMOUD P.
FT Hafner WA, Menoud P;
FT WPI; 2002-148020/19.
FT
PT Formulating molecular probes, useful for diagnosis and therapy, by
XX analyzing the forms of presentation of targets in organisms
XX
PS Disclosure; Fig 1; 25pp; French.
XX
CC The specification describes a process for formulating and selecting
CC non-contagious molecular probes which are used for detecting nucleic
CC acids. The process that takes account of the forms of presentation of
CC cellular organisms during the life cycle, or where these forms depend
CC on different chemical, biological and physical states of the organism.
CC The process is a contamination free diagnostic method for direct
CC quantitative and qualitative analysis and gene typing of infectious
CC agents without prior DNA or RNA extraction. The probes enable detection
CC of nucleic acids in sections of living or immobilised, frozen or fixed
CC tissues. Probes of the invention are useful for detecting viral
CC (including oncogenic), bacterial, animal and plant nucleic acid,
CC e.g. in situ hybridization for detecting presence of (pathogenic)
CC microbes and/or determination of genotype, as primers for polymerase
CC chain reaction amplification and therapeutically for blocking
CC replication of pathogens e.g. as ribozymes. Particularly, the probes
CC are used to detect retroviral particles circulating in body fluids or
CC present in cells. The present sequence represents a sequence comprising
CC a consensus endodogmatic sequence/consensus exodogmatic sequence.
XX
SQ Sequence 209 BP; 48 A; 59 C; 64 G; 38 U; 0 other;

```

```

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAAGCACCCTATCAGCAGTACACAA 28
Db 78 GCAAGCACCCTATCAGCAGTACACAA 51
RESULT 7
ID AAV70460/C
AC AAV70460;
XX
AC AAV70460;
XX
DT 08-APR-1999 (first entry)
XX
DE Partial sequence of HCV subtype 1b amplicon #86.
XX
KW Nucleic acid detection; nucleic acid characterisation; hybridisation;
XX infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.
XX Hepatitis C virus.
XX
PN WO9850403-A1.
XX
PD 12-NOV-1998.
XX
PF 05-MAY-1998; 98WO-US03194.
XX
PR 03-MAR-1998; 98US-0034205.
XX 05-MAY-1997; 97US-0851588.
XX 19-SEP-1997; 97US-0934097.
XX
PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
XX
PI Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;
XX Lyamchev VI, Neel BP, Prudent JR;
XX WPI; 1998-610317/51.
XX
PT Detection and characterisation of nucleic acid sequences - by mixing
PT a folded target and one or more probes to form a probe/folded target
PT complex and detecting and characterising the complexes
XX
PS Example 5; Page 172-173; 279pp; English.
XX
CC The invention relates to methods and compositions of detection and
CC characterisation of nucleic acid sequences and sequence changes. One
CC method of detection and characterisation comprises: (a) providing: (1) a
CC folded target having a DNA sequence comprising at least 1 double
CC 1 probe complementary to at least 1 single stranded region; and (ii) at least
CC (b) mixing the target and probes so that the probe hybridises to form a
CC of structure formation in nucleic acid targets; for analysing folded
CC nucleic acids targets; and for analysis of nucleic acid structures. The
CC methods can be used for the detection and characterisation of nucleic
CC acid sequences indicative of an infection, the presence of pathogenic nucleic acid
CC sequences indicative of an infection, the presence of variants or alleles
CC of mammalian genes associated with disease and cancers, and the
CC as well as in paternity determinations. The methods allow simultaneous
CC analysis of both strands (e.g. the sense and antisense strands) and are
CC ideal for high-level multiplexing. The products produced are amenable to
CC performed in solution or in the solid phase (e.g. on a solid support).
CC The methods are powerful in that they allow for analysis of longer
CC fragments of nucleic acid than current methodologies. Sequences
CC AAV70453-61 represent partial sequences of different amplicons of
CC hepatitis C virus (HCV) subtypes 1a, 1b, 2c and 3a. These partial
XX sequences are used for identifying the HCV subtypes.
XX
SQ Sequence 232 BP; 40 A; 65 C; 76 G; 51 T; 0 other;

```

Query Match 100.0%; Score 28; DB 19; Length 232;
 Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCCCTATCAGCAGTACCAACAA 28
 |||||||
 DB 232 GCAAGCACCCCTATCAGCAGTACCAACAA 205

RESULT 8
 ABL46070/c
 ID ABL46070 standard; DNA; 232 BP.

AC ABL46070;

DT 26-APR-2002 (first entry)

DE Hepatitis C virus partial sequence #86 SEQ ID NO:37.

KW Nucleic acid accessible hybridisation site; detection; hybridisation;
 KW Characterisation; identification; nucleic acid structure; diagnosis;
 KW PCR primer; probe; ss.

OS Hepatitis C virus.
 OS Synthetic.

PN WO200198537-A2.

PD 27-DEC-2001.

PF 15-JUN-2001; 2001WO-US19401.

PR 17-JUN-2000; 2000US-212308P.

PR 15-JUN-2001; 2001US-0212308.

PA (THIR-) THIRD WAVE TECHNOLOGIES INC.

PI Lyemichiev V, Allawi H, Dong F, Neri BP, Vener IT;

DR WPI; 2002-049698/06.

XX Identifying oligonucleotides hybridizing to nucleic acids containing
 PT secondary structure, useful in clinical diagnosis, comprises
 PT identifying primers that interact with the target to form an extension
 PT product under amplification conditions -

PS Example 5; Page 366; 409pp; English.

XX The present invention describes a method for identifying oligonucleotides
 CC with desired hybridisation properties to nucleic acid targets containing
 CC secondary structure. The method comprises amplifying a target nucleic
 CC acid having at least one accessible and one inaccessible site. Primers
 CC that form an extension product are identified as the oligonucleotides
 CC which can interact with the folded target nucleic acid. Oligonucleotides
 CC from the present invention can be used in novel detection methods for
 CC clinical diagnostic purposes, including the detection and identification
 CC of pathogenic organisms (e.g. HIV). The method allows the ability to
 CC rapidly analyse nucleic acid structures. ABL46034 to ABL46367 represent
 CC sequences used in the exemplification of the present invention.

XX Sequence 232 BP; 40 A; 65 C; 76 G; 51 T; 0 other;

Query Match 100.0%; Score 28; DB 24; Length 232;
 Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCCCTATCAGCAGTACCAACAA 28
 |||||||
 DB 232 GCAAGCACCCCTATCAGCAGTACCAACAA 205

RESULT 9
 AAV70455/c

ID AAV70455 standard; DNA; 239 BP.

AC AAV70455;

DT 08-APR-1999 (first entry)

DE Partial sequence of HCV subtype 1a amplicon #72.

KW Nucleic acid detection; nucleic acid characterisation; hybridisation;
 KW Infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.

OS Hepatitis C virus.

PN WO9850403-A1.

PD 12-NOV-1998.

PF 05-MAY-1998; 98WO-US03194.

PR 03-MAR-1998; 98US-0034205.

PR 05-MAY-1997; 97US-0851588.

PR 19-SEP-1997; 97US-0934097.

PA (THIR-) THIRD WAVE TECHNOLOGIES INC.

PI Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L,
 PI Lyemichiev VI, Neri BP, Prudent JR;

DR WPI; 1998-610317/51.

XX Detection and characterisation of nucleic acid sequences - by mixing
 PT a folded target and one or more probes to form a probe/folded target
 PT complex and detecting and characterising the complexes

PS Example 5; Page 171; 279pp; English.

XX The invention relates to methods and compositions of detection and
 CC characterisation of nucleic acid sequences and sequence changes. One
 CC method of detection and characterisation comprises: (a) providing: (1) a
 CC folded target having a DNA sequence comprising at least 1 double
 CC stranded region and at least 1 single stranded region; and (11) at least
 CC 1 probe complementary to at least a portion of the folded target; and
 CC (b) mixing the target and probes so that the probe hybridises to form a
 CC probe/folded target complex. Also provided are methods for determination
 CC of structure formation in nucleic acid targets; for analysing folded
 CC nucleic acids targets; and for analysis of nucleic acid structures. The
 CC methods can be used for the detection and characterisation of nucleic
 CC acid sequences indicative of an infection, the presence of pathogenic nucleic acid
 CC sequences indicative of an infection, the presence of variants or alleles
 CC of mammalian genes associated with disease and cancers, and the
 CC identification of the source of nucleic acids found in forensic samples,
 CC as well as in paternally determinations. The methods allow simultaneous
 CC analysis of both strands (e.g. the sense and antisense strands) and are
 CC ideal for high-level multiplexing. The products produced are amenable to
 CC qualitative, quantitative and positional analysis. The methods may be
 CC performed in solution or in the solid phase (e.g. on a solid support).
 CC The methods are powerful in that they allow for analysis of longer
 CC fragments of nucleic acid than current methodologies. Sequences
 CC AAV70453-61 represent partial sequences of different amplicons of
 CC hepatitis C virus (HCV) subtypes 1a, 1b, 2c and 3a. These partial
 CC sequences are used for identifying the HCV subtypes.

XX Sequence 239 BP; 45 A; 63 C; 78 G; 53 T; 0 other;

Query Match 100.0%; Score 28; DB 19; Length 239;
 Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCCCTATCAGCAGTACCAACAA 28
 |||||||
 DB 239 GCAAGCACCCCTATCAGCAGTACCAACAA 212

RESULT 10
AAV70459/c standard; DNA: 239 BP.
XX AAV70459;
XX
XX AAV70459;
XX
XX 08-APR-1999 (first entry)
XX
XX Partial sequence of HCV subtype 1a amplicon #85.
XX
XX Nucleic acid detection; nucleic acid characterisation; hybridisation;
XX infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.
XX
XX Hepatitis C virus.
XX
XX MO9850403-A1.
XX
XX 12-NOV-1998.
XX
XX 05-MAY-1998; 98MO-US03194.
XX
XX 03-MAR-1998; 98US-0034205.
XX 05-MAY-1997; 97US-0851588.
XX 19-SEP-1997; 97US-0934097.
XX
XX (THIR-) THIRD WAVE TECHNOLOGIES INC.
XX
XX Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;
XX Lyamichev VI, Neri BP, Prudent JR;
XX
XX WPI; 1998-610317/51.
XX
XX Detection and characterisation of nucleic acid sequences - by mixing
XX a folded target and one or more probes to form a probe/folded target
XX complex and detecting and characterising the complexes
XX
XX Example 5; Page 172; 279pp; English.
XX
XX The invention relates to methods and compositions of detection and
XX characterisation of nucleic acid sequences and sequence changes. One
XX method of detection and characterisation comprises: (a) providing: (1) a
XX folded target having a DNA sequence comprising at least 1 double
XX stranded region and at least 1 single stranded region; and (11) at least
XX 1 probe complementary to at least a portion of the folded target; and
XX (b) mixing the target and probes so that the probe hybridises to form a
XX probe/folded target complex. Also provided are methods for determination
XX of structure formation in nucleic acid targets; for analysing folded
XX nucleic acids targets; and for analysis of nucleic acid structures. The
XX methods can be used for the detection and characterisation of nucleic
XX acid sequences to detect the presence of pathogenic nucleic acid
XX sequences indicative of an infection, the presence of variants or alleles
XX of mammalian genes associated with disease and cancers, and the
XX identification of the source of nucleic acids found in forensic samples,
XX as well as in paternity determinations. The methods allow simultaneous
XX analysis of both strands (e.g. the sense and antisense strands) and are
XX ideal for high-level multiplexing. The products produced are amenable to
XX qualitative, quantitative and positional analysis. The methods may be
XX performed in solution or in the solid phase (e.g. on a solid support).
XX The methods are powerful in that they allow for analysis of longer
XX fragments of nucleic acid than current methodologies. Sequences
XX AAV70453-61 represent partial sequences of different amplicons of
XX hepatitis C virus (HCV) subtypes 1a, 1b, 2c and 3a. These partial
XX sequences are used for identifying the HCV subtypes.
XX
XX Sequence 239 BP; 45 A; 64 C; 77 G; 53 T; 0 other;
XX
XX Query Match 100.0%; Score 28; DB 19; Length 239;
XX Best Local Similarity 100.0%; Pred. No. 0.0039;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11
ABL46065/c standard; DNA: 239 BP.
XX ABL46065;
XX
XX ABL46065;
XX
XX 26-APR-2002 (first entry)
XX
XX Hepatitis C virus partial sequence #72 SEQ ID NO:32.
XX
XX Nucleic acid accessible hybridisation site; detection; hybridisation;
XX characterisation; identification; nucleic acid structure; diagnosis;
XX PCR primer; probe; ss.
XX
XX Hepatitis C virus.
XX
XX Synthetic.
XX
XX WO200198537-A2.
XX
XX 27-DEC-2001.
XX
XX 15-JUN-2001; 2001WO-US19401.
XX
XX 17-JUN-2000; 2000US-212308P.
XX 15-JUN-2001; 2001US-0212308.
XX
XX (THIR-) THIRD WAVE TECHNOLOGIES INC.
XX
XX Lyamichev V, Allawi H, Dong F, Neri BP, Vener IT;
XX
XX WPI; 2002-049698/06.
XX
XX Identifying oligonucleotides hybridizing to nucleic acids containing
XX secondary structure, useful in clinical diagnosis, comprises
XX identifying primers that interact with the target to form an extension
XX product under amplification conditions -
XX
XX Example 5; Page 365; 409pp; English.
XX
XX The present invention describes a method for identifying oligonucleotides
XX with desired hybridisation properties to nucleic acid targets containing
XX secondary structure. The method comprises amplifying a target nucleic
XX acid having at least one accessible and one inaccessible site. Primers
XX that form an extension product are identified as the oligonucleotides
XX which can interact with the folded target nucleic acid. Oligonucleotides
XX from the present invention can be used in novel detection methods for
XX clinical diagnostic purposes, including the detection and identification
XX of pathogenic organisms (e.g. HIV). The method allows the ability to
XX rapidly analyse nucleic acid structures. ABL46034 to ABL46367 represent
XX sequences used in the exemplification of the present invention.
XX
XX Sequence 239 BP; 45 A; 63 C; 78 G; 53 T; 0 other;
XX
XX Query Match 100.0%; Score 28; DB 24; Length 239;
XX Best Local Similarity 100.0%; Pred. No. 0.0039;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12
ABL46069/c standard; DNA: 239 BP.
XX ABL46069;
XX
XX ABL46069;
XX
XX 26-APR-2002 (first entry)
XX
XX Hepatitis C virus partial sequence #85 SEQ ID NO:36.
XX

XX Nucleic acid accessible hybridisation site; detection; hybridisation;
 KW characterisation; identification; nucleic acid structure; diagnosis;
 KM PCR primer; probe; ss.
 XX
 XX Hepatitis C virus.
 OS Synthetic.
 XX
 PN WO200198537-A2.
 PD 27-DEC-2001.
 XX
 PE 15-JUN-2001; 2001WO-US19401.
 XX
 PR 17-JUN-2000; 2000US-212308P.
 PR 15-JUN-2001; 2001US-0212308.
 XX
 PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
 XX
 PI Lyamichev V, Allawi H, Dong F, Neri BP, Vener IT;
 XX
 DR WPI, 2002-049698/06.
 XX
 XX Identifying oligonucleotides hybridizing to nucleic acids containing
 PT secondary structure, useful in clinical diagnosis, comprises
 PT identifying primers that interact with the target to form an extension
 PT product under amplification conditions -
 XX
 PS Example 5; Page 366; 409pp; English.
 XX
 CC The present invention describes a method for identifying oligonucleotides
 CC with desired hybridisation properties to nucleic acid targets containing
 CC secondary structure. The method comprises amplifying a target nucleic
 CC acid having at least one accessible and one inaccessible site. Primers
 CC that form an extension product are identified as the oligonucleotides
 CC which can interact with the folded target nucleic acid. Oligonucleotides
 CC from the present invention can be used in novel detection methods for
 CC clinical diagnostic purposes, including the detection and identification
 CC of pathogenic organisms (e.g. HIV). The method allows the ability to
 CC rapidly analyse nucleic acid structures. ABL46034 to ABL46367 represent
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 239 BP; 45 A; 64 C; 77 G; 53 T; 0 other;
 XX

Query Match 100.0%; Score 28; DB 24; Length 239;
 Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCACGACCCCTATCAGGAGTACACAA 28
 ||||||||||||||||||||||||||||
 DB 239 GCACGACCCCTATCAGGAGTACACAA 212

RESULT 13
 AAV70457/C
 ID AAV70457 standard; DNA; 240 BP.
 XX
 AC AAV70457;
 XX
 DT 08-APR-1999 (first entry)
 XX
 DE Partial sequence of HCV subtype 2b amplicon #74.
 XX
 KM Nucleic acid detection; nucleic acid characterisation; hybridisation;
 KM infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9850403-A1.
 XX
 PD 12-NOV-1998.
 PD 05-MAY-1998; 98WO-US03194.
 XX
 PF

XX
 PR 03-MAR-1998; 98US-0034205.
 PR 05-MAY-1997; 97US-0851588.
 PR 19-SEP-1997; 97US-0934097.
 XX
 XX (THIR-) THIRD WAVE TECHNOLOGIES INC.
 PA
 PI Anderson TR, Brow MAD, Dahlberg JE, Dong F, Fors L;
 PI Lyamichev VI, Neri BP, Prudent JR;
 XX
 DR WPI, 1998-610317/51.
 XX
 PT Detection and characterisation of nucleic acid sequences - by mixing
 PT a folded target and one or more probes to form a probe/folded target
 PT complex and detecting and characterising the complexes
 XX
 PS Example 5; Page 171-172; 279pp; English.
 XX
 CC The invention relates to methods and compositions of detection and
 CC characterisation of nucleic acid sequences and sequence changes. One
 CC method of detection and characterisation comprises: (a) providing: (1) a
 CC folded target having a DNA sequence comprising at least 1 double
 CC stranded region and at least 1 single stranded region; and (1i) at least
 CC 1 probe complementary to at least a portion of the folded target; and
 CC (b) mixing the target and probes so that the probe hybridises to form a
 CC probe/folded target complex. Also provided are methods for determination
 CC of structure formation in nucleic acid targets; for analysing folded
 CC nucleic acids targets; and for analysis of nucleic acid structures. The
 CC methods can be used for the detection and characterisation of nucleic
 CC acid sequences indicative of an infection, the presence of variants or alleles
 CC of mammalian genes associated with disease and cancers, and the
 CC identification of the source of nucleic acids found in forensic samples,
 CC as well as in paternity determinations. The methods allow simultaneous
 CC analysis of both strands (e.g. the sense and antisense strands) and are
 CC ideal for high-level multiplexing. The products produced are amenable to
 CC qualitative, quantitative and positional analysis. The methods may be
 CC performed in solution or in the solid phase (e.g. on a solid support).
 CC The methods are powerful in that they allow for analysis of longer
 CC fragments of nucleic acid than current methodologies. Sequences
 CC AAV70453-61 represent partial sequences of different amplicons of
 CC hepatitis C virus (HCV) subtypes 1a, 1b, 2c and 3a. These partial
 CC sequences are used for identifying the HCV subtypes.
 XX
 SQ Sequence 240 BP; 47 A; 64 C; 74 G; 55 T; 0 other;
 XX

Query Match 100.0%; Score 28; DB 19; Length 240;
 Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCACGACCCCTATCAGGAGTACACAA 28
 ||||||||||||||||||||||||||||
 DB 240 GCACGACCCCTATCAGGAGTACACAA 213

RESULT 14
 AAV70456/C
 ID AAV70456 standard; DNA; 240 BP.
 XX
 AC AAV70456;
 XX
 DT 08-APR-1999 (first entry)
 XX
 DE Partial sequence of HCV subtype 1a amplicon #73.
 XX
 KM Nucleic acid detection; nucleic acid characterisation; hybridisation;
 KM infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9850403-A1.
 XX
 PD 12-NOV-1998.
 PD
 XX

XX 05-MAY-1998; 98WO-US03194.
 PF 03-MAR-1998; 98US-0034205.
 PR 05-MAY-1997; 97US-0851588.
 PR 19-SEP-1997; 97US-0934097.
 XX
 PA (THIRD WAVE TECHNOLOGIES INC.
 PI Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;
 PI Lyamchev VI, Neri BP, Prudent JR;
 DR WPI; 1998-610317/51.
 XX
 PT Detection and characterisation of nucleic acid sequences - by mixing
 PT a folded target and one or more probes to form a probe/folded target
 PT complex and detecting and characterising the complexes
 XX
 PS Example 5; Page 171; 279pp; English.

CC The invention relates to methods and compositions of detection and
 CC characterisation of nucleic acid sequences and sequence changes. One
 CC method of detection and characterisation comprises: (a) providing: (1) a
 CC folded target having a DNA sequence comprising at least 1 double
 CC stranded region and at least 1 single stranded region; and (11) at least
 CC 1 probe complementary to at least a portion of the folded target; and
 CC (b) mixing the target and probes so that the probe hybridises to form a
 CC probe/folded target complex. Also provided are methods for determination
 CC of structure formation in nucleic acid targets; for analysing folded
 CC nucleic acids targets; and for analysis of nucleic acid structures. The
 CC methods can be used for the detection and characterisation of nucleic
 CC acid sequences to detect the presence of pathogenic nucleic acid
 CC sequences indicative of an infection, the presence of variants or alleles
 CC of mammalian genes associated with disease and cancers, and the
 CC identification of the source of nucleic acids found in forensic samples,
 CC as well as in paternity determinations. The methods allow simultaneous
 CC analysis of both strands (e.g. the sense and antisense strands) and are
 CC ideal for high-level multiplexing. The products produced are amenable to
 CC qualitative, quantitative and positional analysis. The methods may be
 CC performed in solution or in the solid phase (e.g. on a solid support).
 CC The methods are powerful in that they allow for analysis of longer
 CC fragments of nucleic acid than current methodologies. Sequences
 CC AAU70453-61 represent partial sequences of different amplicons of
 CC hepatitis C virus (HCV) subtypes 1a, 1b, 2c and 3a. These partial
 CC sequences are used for identifying the HCV subtypes.
 XX
 SQ Sequence 240 BP; 47 A; 63 C; 78 G; 52 T; 0 other;

Query Match 100.0%; Score 28; DB 19; Length 240;
 Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGACACCTATCAGGACGATACCAAA 28
 DB 240 GCAGACACCTATCAGGACGATACCAAA 213

RESULT 15
 ID AAU70458/C standard; DNA; 240 BP.
 XX AAU70458;
 AC AAU70458;
 XX
 DT 08-APR-1999 (first entry)
 XX
 DE Partial sequence of HCV subtype 3a amplicon #81.
 XX
 KW Nucleic acid detection; nucleic acid characterisation; hybridisation;
 XX infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.
 OS Hepatitis C virus.
 XX
 PN WC9850403-A1.

XX 12-NOV-1998.
 PD 05-MAY-1998; 98WO-US03194.
 PF 03-MAR-1998; 98US-0034205.
 PR 05-MAY-1997; 97US-0851588.
 PR 19-SEP-1997; 97US-0934097.
 XX
 PA (THIRD WAVE TECHNOLOGIES INC.
 PI Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;
 PI Lyamchev VI, Neri BP, Prudent JR;
 DR WPI; 1998-610317/51.
 XX
 PT Detection and characterisation of nucleic acid sequences - by mixing
 PT a folded target and one or more probes to form a probe/folded target
 PT complex and detecting and characterising the complexes
 XX
 PS Example 5; Page 172; 279pp; English.

CC The invention relates to methods and compositions of detection and
 CC characterisation of nucleic acid sequences and sequence changes. One
 CC method of detection and characterisation comprises: (a) providing: (1) a
 CC folded target having a DNA sequence comprising at least 1 double
 CC stranded region and at least 1 single stranded region; and (11) at least
 CC 1 probe complementary to at least a portion of the folded target; and
 CC (b) mixing the target and probes so that the probe hybridises to form a
 CC probe/folded target complex. Also provided are methods for determination
 CC of structure formation in nucleic acid targets; for analysing folded
 CC nucleic acids targets; and for analysis of nucleic acid structures. The
 CC methods can be used for the detection and characterisation of nucleic
 CC acid sequences to detect the presence of pathogenic nucleic acid
 CC sequences indicative of an infection, the presence of variants or alleles
 CC of mammalian genes associated with disease and cancers, and the
 CC identification of the source of nucleic acids found in forensic samples,
 CC as well as in paternity determinations. The methods allow simultaneous
 CC analysis of both strands (e.g. the sense and antisense strands) and are
 CC ideal for high-level multiplexing. The products produced are amenable to
 CC qualitative, quantitative and positional analysis. The methods may be
 CC performed in solution or in the solid phase (e.g. on a solid support).
 CC The methods are powerful in that they allow for analysis of longer
 CC fragments of nucleic acid than current methodologies. Sequences
 CC AAU70453-61 represent partial sequences of different amplicons of
 CC hepatitis C virus (HCV) subtypes 1a, 1b, 2c and 3a. These partial
 CC sequences are used for identifying the HCV subtypes.
 XX

SQ Sequence 240 BP; 45 A; 66 C; 79 G; 50 T; 0 other;

Query Match 100.0%; Score 28; DB 19; Length 240;
 Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGACACCTATCAGGACGATACCAAA 28
 DB 240 GCAGACACCTATCAGGACGATACCAAA 213

Search completed: July 10, 2003, 19:52:42
 Job time : 60.0831 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 12.5213 seconds

(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631B-7

Perfect score: 28

Sequence: 1 gcaagcaccctatcagcagcagcacaac 28

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	28	2	US-08-738-928-2
2	28	100.0	28	3	US-09-039-866-4
3	28	100.0	47	4	US-09-547-757-1
4	28	100.0	57	1	US-08-356-287-36
5	28	100.0	57	5	PCR-US93-04863-36
6	28	100.0	232	4	US-09-034-205-37
7	28	100.0	232	4	US-08-934-097A-37
8	28	100.0	232	4	US-08-851-588-37
9	28	100.0	232	4	US-09-677-218B-37
10	28	100.0	232	4	US-09-677-218B-37
11	28	100.0	233	4	US-09-034-205-32
12	28	100.0	233	4	US-09-034-205-36
13	28	100.0	233	4	US-08-934-097A-32
14	28	100.0	233	4	US-08-934-097A-36
15	28	100.0	233	4	US-08-851-588-32
16	28	100.0	233	4	US-08-851-588-36
17	28	100.0	233	4	US-09-677-218B-32
18	28	100.0	233	4	US-09-677-218B-36
19	28	100.0	233	4	US-09-677-192-32
20	28	100.0	239	4	US-09-677-192-36
21	28	100.0	240	4	US-09-034-205-33
22	28	100.0	240	4	US-09-034-205-34
23	28	100.0	240	4	US-09-034-205-35
24	28	100.0	240	4	US-08-934-097A-33
25	28	100.0	240	4	US-08-934-097A-34
26	28	100.0	240	4	US-08-934-097A-35
27	28	100.0	240	4	US-08-851-588-33

c 28	28	100.0	240	4	US-08-851-588-34	Sequence 34, Appl
c 29	28	100.0	240	4	US-08-851-588-35	Sequence 35, Appl
c 30	28	100.0	240	4	US-09-677-218B-33	Sequence 33, Appl
c 31	28	100.0	240	4	US-09-677-218B-34	Sequence 34, Appl
c 32	28	100.0	240	4	US-09-677-192-33	Sequence 35, Appl
c 33	28	100.0	240	4	US-09-677-192-34	Sequence 36, Appl
c 34	28	100.0	240	4	US-09-677-192-35	Sequence 37, Appl
c 35	28	100.0	240	4	US-09-677-192-36	Sequence 38, Appl
c 36	28	100.0	244	4	US-09-034-205-26	Sequence 26, Appl
c 37	28	100.0	244	4	US-09-034-205-27	Sequence 27, Appl
c 38	28	100.0	244	4	US-09-034-205-28	Sequence 28, Appl
c 39	28	100.0	244	4	US-09-034-205-29	Sequence 29, Appl
c 40	28	100.0	244	4	US-09-034-205-31	Sequence 31, Appl
c 41	28	100.0	244	4	US-08-934-097A-26	Sequence 26, Appl
c 42	28	100.0	244	4	US-08-934-097A-27	Sequence 27, Appl
c 43	28	100.0	244	4	US-08-934-097A-28	Sequence 28, Appl
c 44	28	100.0	244	4	US-08-934-097A-29	Sequence 29, Appl
c 45	28	100.0	244	4	US-08-934-097A-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-08-738-928-2
Sequence 2, Application US/08738928
Patent No. 5837442

GENERAL INFORMATION:

APPLICANT: Tsang, Sue Y.
TITLE OF INVENTION: Oligonucleotide Primers for Amplifying

TITLE OF INVENTION: HCV Nucleic Acid

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: U.S.A.

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/738,928

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Petry, Douglas A.

REGISTRATION NUMBER: 35,321

REFERENCE/DOCKET NUMBER: 9263

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 814-2974

TELEFAX: (510) 814-2977

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 28 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-738-928-2

Query Match

Best Local Similarity 100.0%, Score 28; DB 2; Length 28;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGCAGCAGCACA 28

DB 1 GCAAGCACCCTATCAGCAGCAGCACA 28

RESULT 2

US-09-039-866-4
Sequence 4, Application US/09039866
Patent No. 6001611

GENERAL INFORMATION:

APPLICANT: Will, Stephen G.
TITLE OF INVENTION: MODIFIED NUCLEIC ACID AMPLIFICATION
TITLE OF INVENTION: PRIMERS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Roche Molecular Systems
STREET: 1080 U.S. Highway 202
CITY: Branchburg
STATE: New Jersey
COUNTRY: United States
ZIP: 08876

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/039,866
FILING DATE:
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Petty, Douglas A.
REGISTRATION NUMBER: 35,321
REFERENCE/DOCKET NUMBER: 1023P
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-09-039-866-4

Query Match
Best Local Similarity 100.0%; Score 28; DB 3; Length 28;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCCTATCAGGAGTCCACAA 28
1 GCAAGCACCCTATCAGGAGTCCACAA 28

Db 1 GCAAGCACCCTATCAGGAGTCCACAA 28

RESULT 3

US-09-547-757-1/c
Sequence 1, Application US/09547757
Patent No. 6368801

GENERAL INFORMATION:

APPLICANT: Faruqi, A. Fawad
TITLE OF INVENTION: Detection and Amplification of RNA using
FILE REFERENCE: Target-Mediated Ligation of DNA by RNA Ligase
CURRENT APPLICATION NUMBER: US/09/547,757
CURRENT FILING DATE: 2000-06-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 47
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic

US-09-547-757-1

Query Match
Best Local Similarity 100.0%; Score 28; DB 4; Length 47;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY

1 GCAAGCACCCTATCAGGAGTCCACAA 28
36 GCAAGCACCCTATCAGGAGTCCACAA 9

Db

RESULT 4

US-08-356-287-36/c
Sequence 36, Application US/08356287
Patent No. 5686272

GENERAL INFORMATION:

APPLICANT: Ronald L. Marshall
APPLICANT: John J. Carrino
TITLE OF INVENTION: AMPLIFICATION OF RNA SEQUENCES USING
TITLE OF INVENTION: THE LIGASE CHAIN REACTION
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy diskette
COMPUTER: Macintosh
OPERATING SYSTEM: System 7.0.1
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,287
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/891,543
FILING DATE: 29 MAY 1992
ATTORNEY/AGENT INFORMATION:
NAME: Paul D. Yassier

REGISTRATION NUMBER: 37,477
REFERENCE/DOCKET NUMBER: 5172, US, P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-938-2623
TELEFAX: 708-937-2341

INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 57
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA

US-08-356-287-36

Query Match
Best Local Similarity 100.0%; Score 28; DB 1; Length 57;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCCTATCAGGAGTCCACAA 28
54 GCAAGCACCCTATCAGGAGTCCACAA 27

Db

RESULT 5

PCT-US93-04863-36/c
Sequence 36, Application PC/TUS9304863

GENERAL INFORMATION:

APPLICANT: Ronald L. Marshall
APPLICANT: John J. Carrino
APPLICANT: Joann C. Sustachek
TITLE OF INVENTION: AMPLIFICATION OF RNA SEQUENCES
TITLE OF INVENTION: USING THE LIGASE CHAIN REACTION
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories

STREET: One Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy diskette
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04863
FILING DATE: 19930524
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/891,543
FILING DATE: 29 MAY 1992
ATTORNEY/AGENT INFORMATION:
NAME: Thomas D. Bralhard
REGISTRATION NUMBER: 32,459
REFERENCE/DOCKET NUMBER: 5172.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-4884
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 57
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA
PCT-US93-04863-36

Query Match 100.0% Score 28; DB 5; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCACGACCCCTATCAGCGATACACAA 28
DB 54 GCACGACCCCTATCAGCGATACACAA 27

RESULT 6
US-09-034-205-37/C
Sequence 37, Application US/09034205
Patent No. 6194149
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Nerli, Bruce P.
TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING
STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,205
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Macknight, Kamrin T.
REGISTRATION NUMBER: 38,230

REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-034-205-37

Query Match 100.0% Score 28; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCACGACCCCTATCAGCGATACACAA 28
DB 232 GCACGACCCCTATCAGCGATACACAA 205

RESULT 7
US-08-934-097A-37/C
Sequence 37, Application US/08934097A
Patent No. 6210880
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Nerli, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
STRUCTURE PROBING WITH STRUCTURE-BRIDGING
OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934,097A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Macknight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-08-934-097A-37

Query Match 100.0% Score 28; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCCTATCAGCAGTACCACAA 28
|||||
Db 232 GCAAGCACCCCTATCAGCAGTACCACAA 205

RESULT 8
US-08-851-588-37/C

; Sequence 37, Application US/08851588
; Patent No. 6214545

GENERAL INFORMATION:

APPLICANT: Doog, Fang
APPLICANT: Lyamichev, Victor I.
APPLICANT: Prudent, James R.
APPLICANT: Dahlberg, James E.
APPLICANT: Fors, Lance
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
TITLE OF INVENTION: Structure Probing
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,588
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02777
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"

US-08-851-588-37

Query Match 100.0%; Score 28; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCCTATCAGCAGTACCACAA 28
|||||
Db 232 GCAAGCACCCCTATCAGCAGTACCACAA 205

RESULT 9
US-09-677-218B-37/C

; Sequence 37, Application US/09677218B
; Patent No. 6355437

GENERAL INFORMATION:

APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Neil, Bruce P.

TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING

STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco.
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/677,218B
FILING DATE: 02-Oct-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/034,205
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"

US-09-677-218B-37

Query Match 100.0%; Score 28; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCCTATCAGCAGTACCACAA 28
|||||
Db 232 GCAAGCACCCCTATCAGCAGTACCACAA 205

RESULT 10

US-09-677-192-37/C
; Sequence 37, Application US/09677192
; Patent No. 6358691

GENERAL INFORMATION:

APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Neil, Bruce P.

TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING STRUCTURE-BRIDGING

FILE REFERENCE: FORS-04708
CURRENT APPLICATION NUMBER: US/09/677,192
CURRENT FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 09/034,205
PRIOR FILING DATE: 1998-03-03
NUMBER OF SEQ ID NOS: 68
SOFTWARE: Patentn Ver. 2.0
SEQ ID NO 37
LENGTH: 232
TYPE: DNA
ORGANISM: Hepatitis C virus
US-09-677-192-37

Query Match 100.0%; Score 28; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Fri Jul 11 15:29:37 2003

QY 1 GCAAGCACCCCTATCAGGAGTACCA 28
Db 232 GCAAGCACCCCTATCAGGAGTACCA 205

RESULT 11
US-09-034-205-32/c
Sequence 32, Application US/09034205
Patent No. 6194149
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Neel, Bruce P.
TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING
STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,205
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamlin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 239 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-034-205-32

Query Match 100.0%; Score 28; DB 4; Length 239;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCCTATCAGGAGTACCA 28
Db 239 GCAAGCACCCCTATCAGGAGTACCA 212

RESULT 12
US-09-034-205-36/c
Sequence 36, Application US/09034205
Patent No. 6194149
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Neel, Bruce P.
TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING
STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,205
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamlin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 239 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-034-205-36

Query Match 100.0%; Score 28; DB 4; Length 239;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCCTATCAGGAGTACCA 28
Db 239 GCAAGCACCCCTATCAGGAGTACCA 212

RESULT 13
US-08-934-097A-32/c
Sequence 32, Application US/08934097A
Patent No. 6210880
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Neel, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
STRUCTURE PROBING WITH STRUCTURE-BRIDGING
OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934,097A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamlin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 239 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

US-08-934-097A-32

Query Match 100.0%; Score 28; DB 4; Length 239;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGACACCTATCAGCAGTACCACAA 28

DB 239 GCAGACACCTATCAGCAGTACCACAA 212

RESULT 14

US-08-934-097A-36/C

Sequence 36, Application US/08934097A

Patent No. 6210880

GENERAL INFORMATION:

APPLICANT: Lyamichev, Victor I.

APPLICANT: Brow, Mary Ann D.

APPLICANT: Fors, Lance

APPLICANT: Nerl, Bruce P.

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid

TITLE OF INVENTION: Structure Probing With Structure-Bridging

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESS: MEDLEN & CARROLL, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/934,097A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Macknight, Kamelin T.

REGISTRATION NUMBER: 38,230

REFERENCE/DOCKET NUMBER: FORS-029980

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 239 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

US-08-934-097A-36

Query Match 100.0%; Score 28; DB 4; Length 239;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGACACCTATCAGCAGTACCACAA 28

DB 239 GCAGACACCTATCAGCAGTACCACAA 212

DB 239 GCAGACACCTATCAGCAGTACCACAA 212

RESULT 15

US-08-851-588-32/C

Sequence 32, Application US/08851588

Patent No. 6214545

GENERAL INFORMATION:

APPLICANT: Dong, Fang

APPLICANT: Lyamichev, Victor I.

APPLICANT: Prudent, James R.

APPLICANT: Dahlberg, James E.

APPLICANT: Fors, Lance

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid

TITLE OF INVENTION: Structure Probing

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESS: MEDLEN & CARROLL, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/851,588

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: FORS-02777

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 239 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

US-08-851-588-32

Query Match 100.0%; Score 28; DB 4; Length 239;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGACACCTATCAGCAGTACCACAA 28

DB 239 GCAGACACCTATCAGCAGTACCACAA 212

Search completed: July 10, 2003, 20:27:39
Job time: 12.5213 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:53:16 ; Search time 76.6382 Seconds
(without alignments)
576.891 Million cell updates/sec

Title: US-10-087-631B-7
Perfect score: 28
Sequence: 1 gcaagcaccctatcagcagcagcaca 28

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/PCRYS_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	28	9	US-10-087-631B-7
2	28	100.0	60	9	US-09-870-939-1
3	28	100.0	232	10	US-09-825-574-37
4	28	100.0	239	10	US-09-825-574-32
5	28	100.0	239	10	US-09-825-574-36
6	28	100.0	240	10	US-09-825-574-33
7	28	100.0	240	10	US-09-825-574-35
8	28	100.0	241	9	US-10-087-631B-10
9	28	100.0	241	9	US-10-087-631B-11
10	28	100.0	242	9	US-10-087-631B-15
11	28	100.0	244	10	US-09-825-574-26
12	28	100.0	244	10	US-09-825-574-27
13	28	100.0	244	10	US-09-825-574-28
14	28	100.0	244	10	US-09-825-574-29
15	28	100.0	244	10	US-09-825-574-31
16	28	100.0	281	9	US-09-940-925A-123
17	28	100.0	281	9	US-09-940-925A-125
18	28	100.0	281	9	US-09-940-925A-126
19	28	100.0	281	9	US-09-940-925A-125

20	28	100.0	281	9	US-09-940-925A-128	Sequence 128, App
21	28	100.0	281	9	US-09-940-925A-129	Sequence 129, App
22	28	100.0	281	9	US-09-940-925A-131	Sequence 131, App
23	28	100.0	281	9	US-09-940-925A-132	Sequence 132, App
24	28	100.0	281	9	US-09-941-193A-123	Sequence 123, App
25	28	100.0	281	9	US-09-941-193A-125	Sequence 125, App
26	28	100.0	281	9	US-09-941-193A-126	Sequence 126, App
27	28	100.0	281	9	US-09-941-193A-128	Sequence 128, App
28	28	100.0	281	9	US-09-941-193A-129	Sequence 129, App
29	28	100.0	281	9	US-09-941-193A-131	Sequence 131, App
30	28	100.0	282	9	US-09-941-193A-132	Sequence 132, App
31	28	100.0	282	9	US-09-940-925A-124	Sequence 124, App
32	28	100.0	282	9	US-09-940-925A-130	Sequence 130, App
33	28	100.0	282	9	US-09-941-193A-124	Sequence 124, App
34	28	100.0	282	9	US-09-941-193A-130	Sequence 130, App
35	28	100.0	286	10	US-09-825-574-21	Sequence 21, App
36	28	100.0	289	10	US-09-825-574-20	Sequence 20, App
37	28	100.0	289	10	US-09-825-574-22	Sequence 22, App
38	28	100.0	289	10	US-09-825-574-23	Sequence 23, App
39	28	100.0	298	10	US-09-345-761-7	Sequence 7, App
40	28	100.0	315	10	US-09-345-761-6	Sequence 6, App
41	28	100.0	341	9	US-09-814-357-3	Sequence 3, App
42	28	100.0	341	9	US-10-255-275-35	Sequence 35, App
43	28	100.0	347	9	US-09-814-352-14	Sequence 14, App
44	28	100.0	347	9	US-10-132-295-1	Sequence 1, App
45	28	100.0	366	9	US-09-992-160-48	Sequence 48, App

ALIGNMENTS

```

RESULT 1
US-10-087-631B-7
Sequence 7, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087, 631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 28
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Description of Artificial Sequence: sr778 HCV-specific primer
US-10-087-631B-7

Query Match      100.0%; Score 28; DB 9; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      1 GCAAGCACCCTATCAGCGAGTACCACAA 28
Db      1 GCAAGCACCCTATCAGCGAGTACCACAA 28

RESULT 2
US-09-870-939-1/C
Sequence 1, Application US/09870939
Publication No. US20020192650A1
GENERAL INFORMATION:
APPLICANT: AMORESE, DOUGLAS A.
APPLICANT: SHANNON, KAREN W.
APPLICANT: COLLINS, PATRICK J.
APPLICANT: WOLBER, PAUL K.
TITLE OF INVENTION: COMPOSITE ARRAYS
FILE REFERENCE: 10010791-1

```

us-10-087-631b-7.rnpb

1 GCACGACCCCTATCAGGCGAGTACCACAA 28
232 GCAAGCACCCCTATCAGGCGAGTACCACAA 29

US-09-825-574-32/c
Sequence 32, Application US/09825574
Patent No. US20020119654A1
GENERAL INFORMATION:
APPLICANT

NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS: Structure Probing with Nucleic Acid
oligonucleotides.
ADDRESSEE: Structure-Bridging
STREET: MEDLEN & CO.

STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE:

COMPUTER: Ibm PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Patent Release #1.0,
APPLICATION NUMBER: US/09/9825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA: <unknown>
APPLICATION NUMBER: 08/934,097
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Macnigh, Kenneth T.
REGISTRATION NUMBER:
REFERENCE:

1. TELECOMMUNICATIONS DOCKET NUMBER: 38, 230
 2. TELEPHONE: (415) 705-8410
 3. INFORMATION FOR FAX: (415) 397-8338
 4. SEQUENCE FOR SEQ ID NO: 3
 5. LENGTH: 239 base pairs
 6. LENTH: 239 base pairs
 7. TYPE: nucleic acid
 8. STRANDEDNESS: double
 9. TOPOLOGY: linear
 10. MOLECULE TYPE: other nucleic acid
 11. DESCRIPTION: /descriptive text/

```

002-574-32 DESCRIPTION: SEQ ID NO: 32
Query Match
Best Local Similarity 100.0%; Score 28;
Matches 28; Conservative 0; Mismatch
1 GCAGCACCCTATCAGCAGTACCA
239 GCA

```

RESULT 5
US-09-825-574-36/c
Sequence 36
Patent No. US2002011945A1
GENERAL INFORMATION
APP. INFORMATION
US-09-825-574-36/c
Sequence 36
Patent No. US2002011945A1
GENERAL INFORMATION
APP. INFORMATION

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.

NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 239 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"

SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-825-574-36

Query Match 100.0%; Score 28; DB 10; Length 239;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAAGCACCTATCAGCAGTACACAA 28
Db 239 GCAAGCACCTATCAGCAGTACACAA 212

RESULT 6
US-09-825-574-33/c
Sequence 33, Application US/09825574
Patent No. US20020119454A1

GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
Brow, Mary Ann D.
Fors, Lance
Neri, Bruce P.

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.

NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"

SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-09-825-574-33

Query Match 100.0%; Score 28; DB 10; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAAGCACCTATCAGCAGTACACAA 28
Db 240 GCAAGCACCTATCAGCAGTACACAA 213

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"

SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-09-825-574-33

Query Match 100.0%; Score 28; DB 10; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAAGCACCTATCAGCAGTACACAA 28
Db 240 GCAAGCACCTATCAGCAGTACACAA 213

RESULT 7
US-09-825-574-34/c
Sequence 34, Application US/09825574
Patent No. US20020119454A1

GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
Brow, Mary Ann D.
Fors, Lance
Neri, Bruce P.

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.

NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-825-574-34

Query Match 100.0%; Score 28; DB 10; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGCAGTACCA 28
DB 240 GCAAGCACCCTATCAGCAGTACCA 213

RESULT 8
US-09-825-574-35/c
Sequence 35, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
Fors, Lance
Neri, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESSES:
ADDRESSEE: MERLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-825-574-35

Query Match 100.0%; Score 28; DB 10; Length 240;

Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGCAGTACCA 28
DB 240 GCAAGCACCCTATCAGCAGTACCA 213

RESULT 9
US-10-087-631B-10/c
Sequence 10, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
TITLE OF INVENTION: CONTROL
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 241
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: DNA sequence derived by
US-10-087-631B-10
OTHER INFORMATION: amplification of HCV type 1 using primers ST280 and ST778

Query Match 100.0%; Score 28; DB 9; Length 241;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGCAGTACCA 28
DB 241 GCAAGCACCCTATCAGCAGTACCA 214

RESULT 10
US-10-087-631B-13/c
Sequence 13, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
TITLE OF INVENTION: CONTROL
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 13
LENGTH: 241
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of artificial sequence: Amplicon sequence derived
OTHER INFORMATION: QS HCV (HCV amplification control having binding sites for ST2
US-10-087-631B-13

Query Match 100.0%; Score 28; DB 9; Length 241;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGCAGTACCA 28
DB 241 GCAAGCACCCTATCAGCAGTACCA 214

RESULT 11
US-10-087-631B-15/c

Sequence 15, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 15
LENGTH: 242
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of artificial sequence: Amplicon derived from ICS62
US-10-087-631B-15

Query Match 100.0%; Score 28; DB 9; Length 242;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGCGAGTACCA 28
DB 242 GCAAGCACCCTATCAGCGAGTACCA 215

RESULT 12
US-09-825-574-26/c
Sequence 26, Application US/09825574
Patent No. US20020119454A1

GENERAL INFORMATION:

APPLICANT: Iyamichev, Victor I.
Brow, Mary Ann D.
Fors, Lance
Neri, Bruce P.

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/825,574

FILING DATE: 03-Apr-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/934,097

FILING DATE: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin T.

REGISTRATION NUMBER: 38,230

REFERENCE/DOCKET NUMBER: FOS-02980

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 244 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-09-825-574-26

Query Match 100.0%; Score 28; DB 10; Length 244;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGCGAGTACCA 28
DB 241 GCAAGCACCCTATCAGCGAGTACCA 214

RESULT 13
US-09-825-574-27/c
Sequence 27, Application US/09825574
Patent No. US20020119454A1

GENERAL INFORMATION:

APPLICANT: Iyamichev, Victor I.
Brow, Mary Ann D.
Fors, Lance
Neri, Bruce P.

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/825,574

FILING DATE: 03-Apr-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/934,097

FILING DATE: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin T.

REGISTRATION NUMBER: 38,230

REFERENCE/DOCKET NUMBER: FOS-02980

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 244 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-09-825-574-27

Query Match 100.0%; Score 28; DB 10; Length 244;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGCGAGTACCA 28
DB 241 GCAAGCACCCTATCAGCGAGTACCA 214

Fri Jul 11 15:29:38 2003

us-10-087-631b-7.rmpb

Page 6

RESULT 14
US-09-825-574-28/c

Sequence 28, Application US/09825574
Patent No. US20020119454A1

GENERAL INFORMATION:

APPLICANT: Lyamichev, Victor I.

Brow, Mary Ann D.

Fors, Lance

Neil, Bruce P.

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/825,574

FILING DATE: 03-Apr-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/934,097

FILING DATE: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin T.

REGISTRATION NUMBER: 38,230

TELECOMMUNICATION INFORMATION:

REFERENCE/DOCKET NUMBER: FORS-02980

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 244 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

SEQUENCE DESCRIPTION: SEQ ID NO: 28:

US-09-825-574-28

Query Match 100.0%; Score 28; DB: 10; Length 244;

Best Local Similarity 100.0%; Pred. No. 0.001; 0; Indels 0; Gaps 0;

Matches 28; Conservative 0; Mismatches 0;

Y 1 GCAAGCACCTATCAGCAGTACCAACA 28

DB 241 GCAAGCACCTATCAGCAGTACCAACA 214

RESULT 15

US-09-825-574-29/c

Sequence 29, Application US/09825574

Patent No. US20020119454A1

GENERAL INFORMATION:

APPLICANT: Lyamichev, Victor I.

Brow, Mary Ann D.

Fors, Lance

Neil, Bruce P.

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/825,574

FILING DATE: 03-Apr-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/934,097

FILING DATE: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin T.

REGISTRATION NUMBER: 38,230

TELECOMMUNICATION INFORMATION:

REFERENCE/DOCKET NUMBER: FORS-02980

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 244 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

SEQUENCE DESCRIPTION: SEQ ID NO: 29:

US-09-825-574-29

Query Match 100.0%; Score 28; DB: 10; Length 244;

Best Local Similarity 100.0%; Pred. No. 0.001; 0; Indels 0; Gaps 0;

Matches 28; Conservative 0; Mismatches 0;

Y 1 GCAAGCACCTATCAGCAGTACCAACA 28

DB 241 GCAAGCACCTATCAGCAGTACCAACA 214

Search completed: July 11, 2003, 15:01:57

Job time : 77.6382 secs

```

FEATURES      location/Qualifiers
              1. .572
              /organism="Silurana tropicalis"

```

/db_xref="taxon:8364"
 /clone="TGAS018013"
 /dev_stage="gastrula (stages 10.5-13 mixed)"
 /lab_host="Escherichia coli XL1-blue"
 /note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5' end of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into PCS107 with EcoRI at the 5' end and NotI at the 3' end."
 BASE COUNT
 ORIGIN
 169 a 198 c 134 g 71 t

Query Match
 Best Local Similarity 85.7%; Pred. No. 54;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCAGACGACCTCTCAGCACCACCA 28
 Db 242 GCAGACGACCTCTCAGCACCACCA 269

RESULT 2
 AL637026 585 bp mRNA linear EST 12-DEC-2001
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 ORGANISM
 AL637026
 AL637026.1 GI:16789005
 EST.
 Western clawed frog.
 Silurana tropicalis
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Silurana.
 1 (bases 1 to 585)
 Huckie E., Taylor R., Ashurst J.L., Zorn A.M. and Rogers J.
 Sanger Xenopus tropicalis EST project 2001 (10_2001)
 Unpublished (2001)
 Contact: Huckie E

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Sanger Centre
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: tropesanger.ac.uk
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS_SEQUENCE.ID: TNeu004n20.sp6
 Sequencing primer: SP6
 This sequence is from a Xenopus Gene Collection (XGC) library
 constructed by Aaron M. Zorn.
 Location/Qualifiers
 1..585

FEATURES
 Source
 /organism="Silurana tropicalis"
 /db_xref="taxon:8364"
 /clone="TNeu004n20"
 /dev_stage="gastrula"
 /lab_host="Escherichia coli DH10B"
 /note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5' end of poly A+ RNA from gastrula. EcoRI-NotI cut cDNA was then ligated into PCS107 with EcoRI at the 5' end and NotI at the 3' end."
 BASE COUNT
 ORIGIN
 171 a 194 c 129 g 91 t

Query Match
 Best Local Similarity 85.7%; Pred. No. 55;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCAGACGACCTCTCAGCACCACCA 28
 Db 32 GCAGACGACCTCTCAGCACCACCA 59

RESULT 3

AL793578 619 bp mRNA linear EST 27-JUN-2002
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 AL793578
 AL793578 XGC-neurula Silurana tropicalis cDNA clone TNeu11121 5',
 mRNA sequence.
 AL793578
 AL793578.1 GI:21579282
 EST.
 Western clawed frog.
 Silurana tropicalis
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Silurana.
 1 (bases 1 to 619)
 Taylor R., Ashurst J.L., Croning M.D.R., Zorn A.M. and Rogers J.
 Sanger Xenopus tropicalis EST project 2002
 Unpublished (2001)
 Contact: Taylor R

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Sanger Centre
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: tropesanger.ac.uk
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS_SEQUENCE.ID: TNeu11121.p1cSP6
 Sequencing primer: P1cSP6
 This sequence is from a Xenopus Gene Collection (XGC) library
 constructed by Aaron M. Zorn.
 Location/Qualifiers
 1..619

FEATURES
 Source
 /organism="Silurana tropicalis"
 /db_xref="taxon:8364"
 /clone="TNeu11121"
 /dev_stage="neurula"
 /lab_host="Escherichia coli DH10B"
 /note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5' end of poly A+ RNA from gastrula. EcoRI-NotI cut cDNA was then ligated into PCS107 with EcoRI at the 5' end and NotI at the 3' end."
 BASE COUNT
 ORIGIN
 182 a 205 c 137 g 95 t

Query Match
 Best Local Similarity 85.7%; Pred. No. 56;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCAGACGACCTCTCAGCACCACCA 28
 Db 61 GCAGACGACCTCTCAGCACCACCA 88

RESULT 4
 B1777787/c 306 bp mRNA linear EST 23-JUL-2002
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 B1777787
 B1777787 EBR008_S0002_B06.R root, 3 week, drought-stressed, cv Optic, EBR008
 Hordeum vulgare cDNA clone EBR008_S0002_B06 5', mRNA sequence.
 B1777787
 B1777787.2 GI:21951028
 EST.
 Hordeum vulgare.
 Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 1 (bases 1 to 306)
 Hedley P., Liu H., Caldwell D., McCallum N., Mudie S., Cardle L.,
 Ramsay L., Machray G., Marshall D.F.M. and Waugh R.
 Development of Barley Transcriptome Resources
 Unpublished (2001)
 On Sep 26, 2001 this sequence version replaced gi:15780679.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Contact: Waugh R, Marshall D.F.
 Genome Dynamics/Computational Biology
 Scottish Crop Research Institute
 Invergowrie, Dundee, DD2 5DA, Scotland, UK
 Tel: 00 44 1382 562731

Query Match 75.7%; Score 21.2; DB 14; Length 731;
Best Local Similarity 88.5%; Pred. No. 89;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCAGCACCCTATCAGGAGTACCAC 26
DB 381 GCAGCTCCGACAGCAGTACCAC 356
|||||

RESULT 7
BHI01607/c
LOCUS
DEFINITION BHI01607 482 bp DNA linear GSS 19-JUN-2001
ACCESSION DNA sequence.
VERSION BHI01607
KEYWORDS BHI01607.1 GI:14926660
SOURCE GSS.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M.,
Tsegaye, G., Geer, K., Kroll, M., Shvartsbeyn, A., Gebregeorgis, E.,
Russell, D., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other GSSs: RPCI-24-229K3.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaot@ig.org

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@email.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end plate: http://www.tigr.org/tldb/Bac_ends/mouse/Bac_end_intro.html
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers
1..482

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-229K3"
/clone_1lb="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAcl; Site 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAcl cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."
BASE COUNT 138 a 87 c 99 g 158 t
ORIGIN

Query Match 73.6%; Score 20.6; DB 17; Length 482;
Best Local Similarity 85.2%; Pred. No. 1,4e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 CAAGCACCCTATCAGGAGTACCACAA 28
DB 462 CAAGCACCCTATCAGGAGTACCACAA 436
|||||

RESULT 8
LOCUS AI657832 560 bp mRNA linear EST 07-JUN-2001
AI657832

DEFINITION

fc23e12.y1 zebrafish Washu MPING EST Danio rerio cDNA clone
IMAGE:3722254.5, similar to SM.YHAF.SCHPO.04360 HYPOTHETICAL 57.6
KD PROTEIN C30D10.15 IN CHROMOSOME II., mRNA sequence.

ACCESSION AI657832
VERSION AI657832.1 GI:4755500
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
Cyprinidae; Danio.
1 (bases 1 to 560)
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy,
S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood,
K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rittler, E.,
Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.,
and Wilson, R.

Washu Zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zdrafish@wustl.edu

CDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 492.

FEATURES

source
Location/Qualifiers
1..560

/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="IMAGE:3722254"
/clone_1lb="zebrafish Washu MPING EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLI-blue MRF"
/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'-GACAGTCTTCTAGATCGGAGCGGCGCCCTTTTCTTTTCTTTT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRU),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRU). Library was constructed
by Matthew Clark (Lehrach lab, ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."

BASE COUNT 162 a 163 c 122 g 113 t
ORIGIN

Query Match 71.4%; Score 20; DB 9; Length 560;
Best Local Similarity 82.1%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GCAGCACCCTATCAGGAGTACCACAA 28
|||||

DB 417 GCACACCCCTGTCGAGTACCAAA 444

RESULT 9
LOCUS B6083128
DEFINITION H3084G10-5 NIA Mouse 15K CDNA Clone Set Mus musculus CDNA clone
ACCESSION B6083128
VERSION B6083128.1 GI:12565696
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka, T.S., Carter, M.G. and Ko, M.S.H.
TITLE Verification and initial annotation of NIA mouse 15K CDNA clone set
JOURNAL Unpublished (2001)
COMMENT Other_ESTS: H3084G10-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit <http://igsun.grc.nia.nih.gov/cdna/15k.html> for details.
Plate: H3084 row: G column: 10
Seq primer: -21M13 Reverse
High quality sequence stop: 595
POLYA-No.

FEATURES
SOURCE Location/Qualifiers
1..595
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:H3084G10-5"
/db_xref="taxon:10090"
/clone="H3084G10"
/clone.lib="NIA Mouse 15K CDNA Clone Set"
/sex="Clones arrayed from a variety of CDNA libraries"
/dev_stage="Clones arrayed from a variety of CDNA libraries"
/lab_host="DH10B"
/note="Vector: pSPOR1; Site_1: SalI; Site_2: NotI. This clone is among a rearrayed set of 15,247 clones from 11 embryo CDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary CDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental CDNA microarray, 2000, Proc. Natl. Acad. Sci. U.S.A. 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

BASE COUNT 143 a 177 c 112 g 163 t

ORIGIN

Query Match 71.4%; Score 20; DB 12; Length 595;
Best Local Similarity 82.1%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCACACCCCTATCAGGACGATCCACAA 28
|||||
DB 501 GCTACACCCCTTAAAGCAGATCCACAA 528

RESULT 10
LOCUS A0117139
DEFINITION A0117139 HEMBA1 Homo sapiens CDNA clone HEMBA1000758 5', mRNA
ACCESSION A0117139
VERSION A0117139.1 GI:10932083
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 808)
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T.
TITLE HRI human CDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1537-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human CDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; CDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES
SOURCE Location/Qualifiers
1..808
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1000758"
/clone.lib="HEMBA1"
/tissue_type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/note="Vector: pME185FL3"

BASE COUNT 213 a 182 c 185 g 225 t

ORIGIN

Query Match 71.4%; Score 20; DB 9; Length 808;
Best Local Similarity 82.1%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCACACCCCTATCAGGACGATCCACAA 28
|||||
DB 669 GCACACCCCTATCAGGACGATCCACAA 696

RESULT 11
LOCUS B6070203/C
DEFINITION H3084G10-3 NIA Mouse 15K CDNA Clone Set Mus musculus CDNA clone
ACCESSION B6070203
VERSION B6070203.1 GI:12552772
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 827)
Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka, T.S., Carter, M.G. and Ko, M.S.H.
TITLE Verification and initial annotation of NIA mouse 15K CDNA clone set
JOURNAL Unpublished (2001)
COMMENT Other_ESTS: H3084G10-5
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA.
 Email: cdna@gsun.grc.nia.nih.gov
 This clone set has been freely distributed to the community. Please
 visit <http://lgsun.grc.nia.nih.gov/cDNA/15k.html> for details.
 Plate: H3084 row: 6 column: 10
 Seq primer: -21M13 Forward
 High quality sequence stop: 827
 POLA-yes.

FEATURES

SOURCE

Location/Qualifiers
 1. 827
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:H3084G10-3"
 /db_xref="taxon:10090"
 /clone="H3084G10"
 /clone_lib="NIA Mouse 15k cDNA Clone Set"
 /sex="Clones arrayed from a variety of cDNA libraries"
 /dev_stage="Clones arrayed from a variety of cDNA libraries"
 /lab_host="DH10B"

/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI. This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryonic ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT) expression profiling of mid-gestation placenta and embryo Proc. Natl. Acad. Sci. U.S.A. 97: 9127-9132; (2) Patterns during preimplantation mouse development, 2000, Development 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and underrepresentation on the X chromosome, 1996, Hum Mol Genet 7: 1967-1978."

BASE COUNT

211 a 184 c 203 g 229 t

ORIGIN

Query Match

Best Local Similarity 82.1%; Score 20; DB 12; Length 827;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

DB 1 GCAGGACCCCTATCAGCAGTACACAA 28
 425 GCTAGCACCCTCTAGCAGATCCACAA 398

RESULT 12
 BG953300 235 bp mRNA linear EST 12-JUN-2001
 LOCUS BG953300
 DEFINITION C4-CT0629-220101-700-909 CT0629 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG953300
 VERSION BG953300.1 GI:14371471
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 235)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

REFERENCE

AUTHORS

TITLE

JOURNAL MEDLINE
 20202663

COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the RAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=C4&t2=C4-CT0629-220101-700-909&t3=2001-01-22&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 124
 High quality sequence stop: 187.
 Location/Qualifiers
 1. 235

FEATURES

SOURCE

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CT0629"
 /dev_stage="Adult"
 /note="Organ: colon; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT

77 a 68 c 47 g 43 t

ORIGIN

Query Match
 Best Local Similarity 70.0%; Score 19.6; DB 13; Length 235;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 DB 3 AACGACGTTATCTGCAGTACACAA 28
 41 AACGACGTTATCTGCAGTACACAA 66

RESULT 13
 BE558358 828 bp mRNA linear EST 23-OCT-2001
 LOCUS BE558358
 DEFINITION HV.CEBD017F16f Hordeum vulgare seedling green leaf EST library
 ACCESSION HVCDBA0005 (Blumeria challenged) Hordeum vulgare cDNA clone
 VERSION BE558358
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 1 (bases 1 to 828)
 Wang, R., Close, T.J., Kleinhofs, A., Wise, R., Wei, F., Begum, D.,
 Fritsch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi,
 D.W., Fenton, R.D., Oates, R. and Main, D.
 Development of a genetically and physically anchored EST resource
 for barley genomics: Blumeria infected incompatible (MI6) seedling
 leaf cDNA library
 unpublished (2001)
 On Aug 14, 2000 this sequence version replaced gi:9822848.
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total hg bases = 113
 Seq primer: AATTACCTCCTCAAGG
 High quality sequence start: 4

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

High quality sequence stop: 517.

FEATURES

SOURCE

1. 828
/organism="Hordeum vulgare"
/cultivar="C116151 (M1a6)"
/db_xref="taxon:4513"
/clone="HV-CEB0017F16f"
/clone.lib="Hordeum vulgare seedling green leaf EST
library HVCDNA0005 (Blumeria challenged)"
/tissue_type="seedling green leaf"
/lab_host="SOLR"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
C.I. 16151 (M1a6) plants were greenhouse grown in the R
Wise lab at Iowa State University, Ames, IA; 7 day old
green seedlings were challenged with isolate 5874 (AVR1a6
of Blumeria graminis f. sp. hordei, and leaves were
harvested 20 and 24 hr post-inoculation and snap frozen;
uninoculated leaves were harvested 20 hr post-inoculation
(Wet, Wise). In the T1 close lab at the University of
California, Riverside, total RNA was prepared from each
sample pool, equal quantities of all three RNA pools were
combined, poly(A) RNA was purified from the mixture, one
primary unamplified cDNA library was made, and 1 million
pfu were in vivo excised to give Bluescript SK(-) cDNA
phagemids (Choi, Close). Phagemids were plated and picked
at the Clemson University Genomics Institute (CUGI) (Begum
Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close T1, Wing R, Kleinholz A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgr/31/cover.html)"

BASE COUNT 200 a 241 c 215 g 171 t 1 others
ORIGIN

Query Match 68.6%; Score 19.2; DB 10; Length 828;
Best Local Similarity 87.5%; Pred No. 6.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 AAGCACCTATCAGCAGTACCAC 26
||||| |
Db 324 AAGCACTCTCTCAGCAGTCCAC 347

RESULT 14
CNS060LX/C 972 bp DNA linear GSS 05-JUL-2001
LOCUS T7 end of clone AM0A007B06 of library AM0A from strain CLIB 89 of
DEFINITION Yarrowia lipolytica, genomic survey sequence.
ACCESSION AL10827
VERSION AL10827.1 GI:12179594
KEYWORDS GSS.
SOURCE Yarrowia lipolytica.
ORGANISM Yarrowia lipolytica.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Dipodomycetes; Yarrowia.
1 (bases 1 to 972)
Soutier,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bollin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
deMontigny,J., Dujon,B., Durans,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

TITLE
JOURNAL

MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 972)
AUTHORS Casaregola,S., Neuvéglise,C., Lepingle,A., Bon,E., Feynereol,C.,
Artiguenave,F., Wincker,P. and Gallardin,C.
TITLE Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia
lipolytica
FEBR Lett. 487 (1), 95-100 (2000)

JOURNAL MEDLINE 20584727
PUBMED 11152892
REFERENCE 3 (bases 1 to 972)
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
SOURCE location/Qualifiers
1. 972
/organism="Yarrowia lipolytica"
/strain="CLIB 89"
/db_xref="taxon:4952"
/clone="AM0A007B06"
/clone.lib="AM0A"
/note="end : 17"
/note="complement<1. .>867)
/note="similar to Saccharomyces cerevisiae ORF YJL034w [

misc_feature
BASE COUNT 208 a 247 c 264 g 232 t 21 others
ORIGIN

Query Match 68.6%; Score 19.2; DB 17; Length 972;
Best Local Similarity 80.8%; Pred No. 7.4e+02;
Matches 21; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 AAGCACCTATCAGCAGTACCACAA 28
||||| |
Db 387 AAGCACTTGTGCACGACGACACAA 362

RESULT 15
A1151927 475 bp mRNA linear EST 30-SEP-1998
LOCUS u147h03.y1 Sugano mouse embryo meva Mus musculus cDNA clone
DEFINITION IMAGE:1885589 5', mRNA sequence.
ACCESSION A1151927
VERSION A1151927.1 GI:3680396
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 475)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project

Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LIND; contact the
 IMAGE Consortium (info@image.lind.gov) for further information.
 MGI:969913

Seq primer: custom primer used
 High quality sequence stop: 450.

FEATURES

source

1..475

/organism="Mus musculus"

/strain="C57BL"

/db_xref="taxon:10090"

/clone="IMAGE:1885589"

/clone_lib="Sugano mouse embryo mewa"

/dev_stage="embryo, 14 dpc"

/lab_host="DH10B"

/note="Vector: pME18S-FL3; site_1: DraIII (CACGTGTG);

site_2: DraIII (CACCATGTG); 1st strand cDNA was primed

with an oligo(dT) primer (ATGCGGCTTTTCTTTTCTTTT);

double-stranded cDNA was ligated to a DraIII adaptor

(TGTGGCCTACTGCG), digested and cloned into distinct DraIII

sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site

CACCATGTG). XhoI should be used to isolate the cDNA

<1.5kb. Library selection was performed to exclude fragments

(University of Tokyo, Institute of Medical Science).

Custom primers for sequencing: 5' end primer

CTTCTGCTCTAAAGTCGCG and 3' end primer

CGACCTGAGCTCGACACA."

BASE COUNT

149 a 80 c 91 g 154 t 1 others

1 others

Query Match

Best Local Similarity 67.9% Score 19; DB 9; Length 475;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCAAGCACCTATCAGGCACTACCA 27

DB 341 GCAAGCACCTATCAGGCACTACCA 367

Search completed: July 11, 2003 02:25:58
 Job time: 511.434 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 225.294 Seconds
(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631b-8

Perfect score: 26
Sequence: 1 cgtcttcgcagatcgctacat 26

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl1:
1: gb_ba:*
2: gb_bt:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_pat:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.2	73.8	139534	2	AP005640
2	18.6	71.5	108284	2	AC122161
3	18.6	71.5	112538	2	AC122171
4	18.6	71.5	145424	2	AC114235
5	18.6	71.5	198868	2	AC110415
6	18.6	71.5	208396	2	AC003101
7	18.6	71.5	212030	2	AC019148
8	17.8	68.5	1290	3	AF358892
9	17.8	68.5	1290	3	AF358893
10	17.8	68.5	1290	3	AF358894
11	17.8	68.5	1290	3	AF358895
12	17.8	68.5	1290	3	AF358897
13	17.8	68.5	1290	3	AF358899
14	17.8	68.5	1290	3	AF359001
15	17.8	68.5	1290	3	AF359003
16	17.8	68.5	1290	3	AF359005
17	17.8	68.5	1679	3	TRBTR
18	17.8	68.5	13051	1	AE009111
19	17.8	68.5	14856	1	AE008077
20	17.8	68.5	20389	1	AE008768
21	17.8	68.5	70100	8	AP001306
22	17.8	68.5	315000	1	RME603644
23	17.6	67.7	1333	8	AY128313
24	17.6	67.7	2191	8	AY039862
25	17.6	67.7	2638	8	AY051704
26	17.6	67.7	3273	8	PMCEFCIT1
27	17.6	67.7	39780	2	AC014011
28	17.6	67.7	54115	2	AC018149
29	17.6	67.7	60040	2	AC005128
30	17.6	67.7	76170	25	ATAC3329
31	17.6	67.7	85660	9	AC092845
32	17.6	67.7	92085	2	AC130938
33	17.6	67.7	95613	2	AC020389
34	17.6	67.7	99725	8	ATF18P4
35	17.6	67.7	101647	8	AC004665
36	17.6	67.7	110726	2	AC131325
37	17.6	67.7	111705	2	AC096875
38	17.6	67.7	117080	2	AC098840
39	17.6	67.7	128903	2	AC093952
40	17.6	67.7	139367	2	AC105676
41	17.6	67.7	151610	3	AC009462
42	17.6	67.7	152021	2	AC010923
43	17.6	67.7	153480	2	AC130879
44	17.6	67.7	155557	3	AC099020
45	17.6	67.7	155840	2	AC006495

ALIGNMENTS

RESULT 1
AP005640
LOCUS
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 2 clone
OSJNBa0006015. *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AP005640.1 GI:22266684
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Sasaki, T., Matsumoto, T. and Katayose, Y.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC
clone:OSJNB0006015
Published Only in Database (2002)
2 (bases 1 to 139534)
Sasaki, T., Matsumoto, T. and Katayose, Y.
Direct Submission
Submitted (15-AUG-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
Tel:81-298-38-7441, Fax:81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1. 139534
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
/clone="OSJNB0006015"
BASE COUNT
40132 a 29480 c 29616 g 40306 t
ORIGIN

Query Match
Best Local Similarity 73.8% Score 19.2; DB 2; Length 139534;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Ory 1 CGTCTTCGAGATCGGTACCTCA 24
DB 58150 CCTCTTCGAGAGCGCTACCTCA 58173

RESULT 2
AC122161/c
LOCUS
DEFINITION
AC122161
VERSION
AC122161.6 GI:21903586
KEYWORDS
HTGS: HTGS.PHASE1; HTGS: DRAFT.
SOURCE
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eutrosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 108284)
Shauli, S., Lin, S., Dixon, R., May, G., Summer, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B. A.
Medicago truncatula BAC Clone mth2-1113
Unpublished
2 (bases 1 to 108284)
Shauli, S., Lin, S., Dixon, R., May, G., Summer, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B. A.
Direct Submission
Submitted (23-MAY-2002) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 108284)
Shauli, S., Lin, S., Dixon, R., May, G., Summer, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B. A.
Direct Submission
Submitted (18-JUL-2002) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,

COMMENT
OK 73019, USA
On Jul 18, 2002 this sequence version replaced gi:21747483.
----- Genome Center
Center: Department of Chemistry And Biochemistry
The University of Oklahoma
Center code:UOKNOR

* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2243 2242: contig of 2242 bp in length
2243 2342: gap of unknown length
2343 5403: contig of 3061 bp in length
5403 5503: gap of unknown length
5503 8006: contig of 2503 bp in length
8006 8107: gap of unknown length
8107 10617: contig of 2511 bp in length
10617 10717: gap of unknown length
10717 13040: contig of 2323 bp in length
13040 13141: gap of unknown length
13141 15966: contig of 2826 bp in length
15966 16067: gap of unknown length
16067 18585: contig of 2518 bp in length
18585 18685: gap of unknown length
18685 21080: contig of 2396 bp in length
21080 21180: gap of unknown length
21180 24824: contig of 3644 bp in length
24824 24924: gap of unknown length
24924 28587: contig of 3663 bp in length
28587 28688: gap of unknown length
28688 31503: contig of 2816 bp in length
31503 31603: gap of unknown length
31603 35512: contig of 3909 bp in length
35512 35613: gap of unknown length
35613 40639: contig of 5027 bp in length
40639 40739: gap of unknown length
40739 47025: contig of 6286 bp in length
47025 47125: gap of unknown length
47125 59249: contig of 12124 bp in length
59249 59350: gap of unknown length
59350 71703: contig of 12354 bp in length
71703 71803: gap of unknown length
71803 82909: contig of 11106 bp in length
82909 82910: gap of unknown length
82910 83010: contig of 25275 bp in length.
Location/Qualifiers
1. 108284
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="mth2-1113"
/clone="mth2-1113"
BASE COUNT
34779 a 17656 c 17963 g 36138 t 1748 others
ORIGIN

Query Match
Best Local Similarity 71.5% Score 18.6; DB 2; Length 108284;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Ory 1 CGTCTTCGAGATCGGTACCTCA 25
DB 27334 CGTCTTCATATACGTACATATA 27310

RESULT 3
AC122171
LOCUS
DEFINITION
AC122171
Medicago truncatula clone mth2-23b18, WORKING DRAFT SEQUENCE, 13
unordered pieces.

ACCESSION	VERSION	KEYWORDS	ORGANISM	SOURCE
AC122171	7	GI:22038532		
HTG: HTGS_PHASE1: HTGS_DRAT.				
Barrel medic.				
Medicago truncatula				
Eumariota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots: Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae				
Medicago.				
1 (bases 1 to 112538)				
Shuaili, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,				
Cook, D., Kim, D. and Roe, B. A.				
Medicago truncatula BAC Clone mth2-23b18				
2 (bases 1 to 112538)				
Shuaili, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,				
Cook, D., Kim, D. and Roe, B. A.				
Direct Submission				
Submitted (23-May-2002) Department of Chemistry And Biochemistry,				
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,				
OK 73019, USA				
On Aug 1, 2002 this sequence version replaced gi:22002170.				

Genome Center				
Center: Department of Chemistry And Biochemistry				
The University of Oklahoma				
Center code:UOKNOR				

* NOTE: This is a 'working draft' sequence. It currently				
* consists of 13 contigs. The true order of the pieces				
* is not known and their order in this sequence record is				
* arbitrary. Gaps between the contigs are represented as				
* runs of N, but the exact sizes of the gaps are unknown.				
* This record will be updated with the finished sequence				
* as soon as it is available and the accession number will				
* be preserved.				
1				
2062: contig of 2062 bp in length				
2063				
2162: gap of unknown length				
2163				
5510: contig of 3348 bp in length				
5511				
5610: gap of unknown length				
5611				
8273: contig of 2663 bp in length				
8274				
8373: gap of unknown length				
8374				
11827: contig of 3434 bp in length				
11828				
11927: gap of unknown length				
11928				
16497: contig of 4570 bp in length				
16498				
16597: gap of unknown length				
16598				
19391: contig of 2794 bp in length				
19392				
19491: gap of unknown length				
19492				
23549: contig of 6058 bp in length				
25550				
25649: gap of unknown length				
34242: contig of 8533 bp in length				
34243				
34342: gap of unknown length				
34343				
41672: contig of 7330 bp in length				
41673				
41772: gap of unknown length				
41773				
50886: contig of 9114 bp in length				
50887				
50986: gap of unknown length				
50987				
63855: contig of 12669 bp in length				
63856				
63955: gap of unknown length				
63956				
86115: contig of 22160 bp in length				
86116				
86215: gap of unknown length				
86216				
112538: contig of 26523 bp in length.				
Location/Qualifiers				
1. 112538				
/organism="Medicago truncatula"				
/db_xref="taxon:3880"				
/clone="mth2-23b18"				
/clone_lib="Medicago truncatula BAC library H2"				

BASE COUNT 36329 a 19038 c 18987 g 36975 t 1209 others

Query Match 71.5%, Score 18.6; DB 2; Length 112538;
Best Local Similarity 84.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 41505 CGTCTTCTCATATCGTACTTAA 41529

RESULT 4
AC114235/c 145424 bp DNA -linear HTG 13-JUL-2002

LOCUS
DEFINITION
Rattus norvegicus clone CH230-334N2, *** SEQUENCING IN PROGRESS
**, 40 unordered pieces.

ACCESSION
AC114235
AC114235 2 GI:21738340

KEYWORDS
HTG; HTGS_PHASEL.

SOURCE
Norway rat.
Rattus norvegicus

ORGANISM
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 145424)
Munzy D.M., Adams C., Adio-Oduola B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbaria, J., Benton, J., Blmage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Bunyah, C., Butch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, D., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
David, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Dem, A.L., Ding, Y., Dinh, H.H.,
Dothwaite, K.J., Draper, H., Dugan-Kocha, S., Durbin, K.J.,
Earhart, T.C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabris, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hughes, M., Holloway, C., Hollins, B.,
Homs, I.F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsoson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kover, C.,
Kratovic, L., Kureshi, A., Landry, R., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Licharge, O., Lileu, C., Liu, J., Liu, W., Lousaged, H.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mel, G., Metker, M.,
Miner, G., Miner, Z., Mitchell, T., Monabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenko, S., Ogun, M., Okwunu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peterson, L., Pickens, R., Primus, E., Pu, L.L., Qulles, M., Ren, Y.,
Rives, M., Rojas, A., Rojubenkan, I., Rolfe, M., Ruiz, S., Savery, G.,
Schier, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Swalek, A., Tabor, P., Tamerlis, A., Tamerlis, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Tellido, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleciyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
Unpublished
Direct Submission
2 (bases 1 to 145424)
Morley, K.C.
Direct Submission
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

Barbata, J., Benton, J., Blincke, K., Blankenburg, K., Bonini, D.,
 Bouck, J., Boyle, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.,
 Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, N., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homsi, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kretovc, J., Kueshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mauney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okunolu, G.,
 Orquenie, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivers, A., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
 Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, C., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, Y., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, J., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 198868)
 Worley, K.C.
 Direct Submission
 Submitted (12-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 198868)
 Worley, K.C.
 Direct Submission
 Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:18851028.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GSRL
 Center clone name: CH230-126C16
 ----- Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 149839 bases at least Q40
 Consensus quality: 156891 bases at least Q30
 Consensus quality: 162963 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 58 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	1081: contig of 1081 bp in length
1082	1181: gap of unknown length
1182	2375: contig of 1194 bp in length
2376	2475: gap of unknown length
2476	3595: contig of 1120 bp in length
3596	3695: gap of unknown length
3696	4903: contig of 1208 bp in length
4904	5003: gap of unknown length
5004	6036: contig of 1033 bp in length
6037	6136: gap of unknown length
6137	7512: contig of 1376 bp in length
7513	7612: gap of unknown length
7613	8924: contig of 1312 bp in length
8925	9024: gap of unknown length
9025	10149: contig of 1125 bp in length
10150	10249: gap of unknown length
10250	11272: contig of 1023 bp in length
11273	11372: gap of unknown length
11373	13108: contig of 1736 bp in length
13109	13208: gap of unknown length
13209	15193: contig of 1985 bp in length
15194	15293: gap of unknown length
15294	16884: contig of 1551 bp in length
16885	16984: gap of unknown length
16985	19379: contig of 2255 bp in length
19380	19379: gap of unknown length
19380	20807: contig of 1428 bp in length
20808	20907: gap of unknown length
20908	23046: contig of 2139 bp in length
23047	23146: gap of unknown length
23147	24465: contig of 1719 bp in length
24466	24965: gap of unknown length
24966	26813: contig of 1848 bp in length
26814	26913: gap of unknown length
26914	29658: contig of 2745 bp in length
29659	29758: gap of unknown length
29759	31694: contig of 1936 bp in length
31695	31794: gap of unknown length
31795	33986: contig of 2192 bp in length
33987	34086: gap of unknown length
34087	36029: contig of 1943 bp in length
36030	36129: gap of unknown length
36130	38429: contig of 2300 bp in length
38430	38529: gap of unknown length
38530	40701: contig of 2172 bp in length
40702	40801: gap of unknown length
40802	42597: contig of 1796 bp in length
42598	42697: gap of unknown length
42698	44891: contig of 2194 bp in length
44892	44991: gap of unknown length
44992	48321: contig of 3330 bp in length
48322	48421: gap of unknown length
48422	50838: contig of 2417 bp in length
50839	50938: gap of unknown length
50939	52492: contig of 1554 bp in length
52493	52592: gap of unknown length
52593	55492: contig of 2900 bp in length
55493	55592: gap of unknown length
55593	57986: contig of 2394 bp in length
57987	58086: gap of unknown length
58087	60471: contig of 2385 bp in length
60472	60571: gap of unknown length
60572	64863: contig of 4292 bp in length
64864	64963: gap of unknown length
64964	67939: contig of 2976 bp in length
67940	68039: gap of unknown length
68040	71114: contig of 3075 bp in length

Qy	2	GTCTTTCGAGATCGGTACTCAAT	26	mismatches	4	Indels	0	Gaps	0
Db	148785	GTCTTTCGAGCTGTACCCCAAT	148809						

RESULT 6	AC003101	LOCUS	AC003101	208396 bp	DNA	linear	PRI 05-JUN-1998
		DEFINITION	Homo sapiens	chromosome 17,	clone	HRPC1C3,	complete sequence.
		ACCESSION	AC003101				
		VERSION	AC003101.1	GI:3184508			
		KEYWORDS	HTG.				
		SOURCE	Homo sapiens				
		ORGANISM	Homo sapiens				

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 208396)
Birren, B., Fasman, K., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 17, clone HRPCC1C23
Unpublished
2 (bases 1 to 208396)
Birren, B., Fasman, K., Linton, L., Nusbaum, C. and Lander, E.

Richardson, P., Lander, E., Ravenen, K., Munro, C., Nussbaum, C., Cooke, P., Daly, M. J., DeParye, E., Devron, K., Dewar, K., Cantu, C., Forrest, C., Gage, D., Gensheimer, S., Gerafsky, K., Gilmarin, T., Hagos, B., Hahnel, I., Harris, K., Howland, J. C., Huang, Y., Jacotot, L., Lincoln, L., Mackenzie, J., Margulis, J., Mull, L., McGarr, A., Mollia, M., Morris, W., Morrow, J., Nachman, A., Naylor, J.

O'Connor, T., Olotu, A., Peterson, K., Roberts, D., Rollins, G., Sernaik, A., Shiu, P., Shyam, R., Stillwell, J., Stone, C., Strickland, C., Sydney, K., Tang, L., Zemseva, I., Zhao, J. and Zody, M.
Direct Submission (13-NOV-1997) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 208356)
Birren, B., Fawcett, D.

TITLE Direct Submission
JOURNAL Submitted (05-JUN-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jun 5, 1998 this sequence version replaced ci:31686331.
 All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>.
FEATURES location/Qualifiers
 source 1..208396

```

Repeat_region      27. 209      repeat de Jong, human PAC library
/rpt_family="MIR"
complement(378. 403)
/rpt_family="(CAT)n"
489. 734
Repeat_region
/rpt_family="AluXs"
1139. 1440
Repeat_region
/rpt_family="AluY"
2566. 2634
/rpt_family="MIR"
complement(2905. 2964)
/rpt_family="MIR"
complement(4148. 4257)
Repeat_region
/rpt_family="MIR"
complement(5399. 5473)
/rpt_family="MIR"
complement(5674. 5787)
/rpt_family="MIR"
complement(6063. 6120)
Repeat_region
/rpt_family="MIR"
complement(6143. 6229)
/rpt_family="MER45"
7875. 8160
Repeat_region
/rpt_family="AluB"
8163. 8465
/rpt_family="AluSq"
complement(8521. 8826)
/rpt_family="AluJb"
complement(8869. 9255)
/rpt_family="L1ME1"
complement(9303. 9607)
/rpt_family="L1M4"
complement(9608. 9907)
/rpt_family="AluY"
complement(9909. 10262)
/rpt_family="L1M4"

```

```

repeat_region complement(10288. .10591)
/rpt_family="Aluuo"
repeat_region complement(10592. .10820)
/rpt_family="L1M4"
repeat_region complement(11163. .11163)
/rpt_family="AluSp"
repeat_region complement(11168. .11260)
/rpt_family="L1M4"
repeat_region complement(11280. .11573)
/rpt_family="AluSc"
repeat_region complement(11615. .11917)
/rpt_family="AluSg"
repeat_region complement(11957. .12251)
/rpt_family="Aluuo"
repeat_region complement(12253. .12681)
/rpt_family="L1M4"
repeat_region complement(12701. .12956)
/rpt_family="L1M4"
repeat_region complement(12976. .13276)
/rpt_family="AluSg"
repeat_region complement(13277. .13417)
/rpt_family="L1M4"
repeat_region complement(13424. .13606)
/rpt_family="Aluud"
repeat_region complement(13657. .13911)
/rpt_family="L1M4"
repeat_region complement(14096. .14403)
/rpt_family="L1MD"
repeat_region complement(14527. .14634)
/rpt_family="AluSc"
repeat_region complement(14635. .14933)
/rpt_family="AluSc"
repeat_region complement(15347. .15347)
/rpt_family="MIR"
repeat_region complement(16322. .16354)
/rpt_family="AluSg"
repeat_region complement(16492. .16492)
/rpt_family="MIR"
repeat_region complement(16657. .16957)
/rpt_family="AluSg"
repeat_region complement(17013. .17317)
/rpt_family="AluSc"
repeat_region complement(18766. .18766)
/rpt_family="L2"
repeat_region complement(19009. .19009)
/rpt_family="FRAM"
repeat_region complement(19035. .19146)
/rpt_family="AluSg"
repeat_region complement(19177. .19309)
/rpt_family="L2"
repeat_region complement(19906. .19906)
/rpt_family="AluSc"
repeat_region complement(20121. .20414)
/rpt_family="L2"
repeat_region complement(20415. .20717)
/rpt_family="Aluud"
repeat_region complement(20719. .20808)
/rpt_family="L2"
repeat_region complement(21025. .21151)
/rpt_family="L2"
repeat_region complement(22356. .22356)
/rpt_family="MER81"
repeat_region complement(22414. .22718)
/rpt_family="Aluud"
repeat_region complement(23218. .23218)
/rpt_family="Aluud"
repeat_region complement(24532. .24532)
/rpt_family="MER21B"
repeat_region complement(24963. .25259)
/rpt_family="Aluuo"
repeat_region complement(25269. .25384)
/rpt_family="MER21B"
repeat_region complement(26217. .26321)
/rpt_family="AluSg"
repeat_region complement(26787. .26834)
/rpt_family="MIR"
repeat_region complement(26835. .27117)
/rpt_family="Aluud"
repeat_region complement(29005. .29005)
/rpt_family="FLAM_C"
repeat_region complement(29274. .29658)
/rpt_family="L1M4"
repeat_region complement(29660. .29977)
/rpt_family="AluSc"
repeat_region complement(30027. .30526)
/rpt_family="L1M4"
repeat_region complement(30536. .30837)
/rpt_family="AluSc"
repeat_region complement(30938. .31065)
/rpt_family="L1M4"
repeat_region complement(31068. .31374)
/rpt_family="AluSg"
repeat_region complement(31576. .31876)
/rpt_family="AluSg"
repeat_region complement(32053. .32347)
/rpt_family="Aluud"
repeat_region complement(32460. .32677)
/rpt_family="MERS8A"
repeat_region complement(32727. .33028)
/rpt_family="AluSc"
repeat_region complement(33679. .33985)
/rpt_family="AluSc"
repeat_region complement(35256. .35582)
/rpt_family="L1M4"
repeat_region complement(35602. .35906)
/rpt_family="AluSg"

```

```

Query Match 71.5% Score 18.6; DB 9; Length 208396;
Best Local Similarity 84.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CGCTTCGACGATGCTACCTCA 25
DB 152067 CCTCTTCACAGATGCTACCTCA 152091

```

```

RESULT 7
LOCUS AC019148 212030 bp DNA linear HMG 19-JUL-2000
DEFINITION Homo sapiens chromosome 17 clone RP11-772K16, WORKING DRAFT
ACCESSION AC019148
VERSION AC019148.4 GI:9230872
KEYWORDS HTG; HMG; PHASE1; HMG; DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 212030)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 212030)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 16, 2000 this sequence version replaced gi:7940422.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0772K16

```

Chemistry: Dye: primer ET; 99% of reads
 Assembly: Dye-terminator Big Dye: 1% of reads
 Consensus quality: 200508 bases at least 40
 Consensus quality: 204858 bases at least 30
 Consensus quality: 207122 bases at least Q30
 Insert size: 256000; agarose-*fp*
 Insert size: 210630; sum-of-*contigs*
 Quality coverage: 4.30 in Q20 bases; agarose-*ft*
 Quality coverage: 4.57 in Q20 bases; *seq*

NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES	source
212030	contig of 55256 bp in length
Location/Qualifiers	
1. 212030	

```
misc_feature 1. 1596
note="assembly_name:Contig4
1697. 3381
note="assembly_name:Contig5
3482. 5702
note="assembly_name:Contig6
5803. 7839
note="assembly_name:Contig7
7940. 11143
note="assembly_name:Contig8"
11244. 14659
note="assembly_name:Contig9"
14760. 19764
note="assembly_name:Contig10
clone_end:sp6
vector_side:left"
```

```

misc_feature      1965.  .28810
/notes="assembly_name:Contig1"
misc_feature      2891.  .42897
/notes="assembly_name:Contig12"
misc_feature      4298.  .59239
/notes="assembly_name:Contig13"
misc_feature      59340. 73503
/notes="assembly_name:Contig14"
misc_feature      73604. 89175
/notes="assembly_name:Contig15"
misc_feature      89276. 118497
/notes="assembly_name:Contig16"
misc_feature      118598. 156674
/notes="assembly_name:Contig17"
misc_feature      156775. 212030
/notes="assembly_name:Contig18"
BASE COUNT      57448 a 48441 c 49876 g 54852 t 1413 others
ORIGIN

```

Query Match	71.5%	Score 18.6;	DB 2;	length 212030;
Best Local Similarity	84.0%;	Pred. No. 1.3+02;		
Matches	21;	Conservative	0;	Mismatches 4;
Indels	0;	Gaps	0;	
QY	1	CCTCTTCGACGATGGCACTCTCA	25	
		1		
DB	101558	CCTCTTCAGATGGCACTCTCA	101582	

RESULT 8	AF358992/c	LOCUS			
DEFINITION	AF358992	1290 bp	DNA	linear	INV 05-JUL-2001
ACCESSION	AF358992	gene, partial cds.	trypanosoma cruzi strain ESMERALDO c13	trypanothione reductase (TR)	
VERSION	AF358992.1	GI:14600021			
KEYWORDS					
SOURCE	Trypanosoma cruzi.				

ORGANISM	REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED
Trypanosoma cruzi	Eukaryota: Euglenozoa: Kinetoplastida: Trypanosomatidae; 1 (bases 1 to 1290)	
Trypanosoma: Schizotrypanum	Machado, C.A. and Ayala, F.J.	
Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of Trypanosoma cruzi	Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)	21310003
		11416213
	2 (bases 1 to 1290)	
Machado, C.A. and Ayala, F.J.		
Direct Submission		
Submitted (12-MAR-2001)	Ecology and Nat.	

FEATURES.	Location/Qualifiers	Location/Qualifiers
5269/-2525, USA	Steinhaus Hall, Irvine, CA	
source	1. .1290	

```

gene      .
          /organism="Trypanosoma cruzi"
          /strain="ESMERALDO c13"
          /db_xref="taxon:5693"
          <1. .>1290
          /gene="tr"
          <1. .>1290
          /gene="tr"
          /product="trypanothione reductase"
          <1. .>1290
          /gene="tr"
          /codon_start=1
          /product="trypanothione reductase"
          /protein_id="AKY1210.1"
          /db_xref="GI:1600022"
          /translation="TLIKRRAVLVDVOMGHPFESALGCTGVNVCYPRKLTAMTGAQ
          GSLEKRVNVRKESADPAVAKERLETEINLLASSVPMMPRIPIETIEHCIELEFFLDM
          PEPKRVLTGVGGFISVEFAGINNAKPKDGGVTLCYGENMILNGFDHITLISNNAFL
          K"
CDS

```

TANGIOILITKENPAKVELNADSKSVTESGKKMDFLVMALIGSPPTKDLQONAG
VMKNGGVQVDEYSRTNYSNITAGDYTNRMPLPVPALNEAALVDYFETNPRKTOH
TRVASAVFSIPPIGTCGLIEEVASKRIVEVAIVYSSFPPLMHNISGSKYKTFVAKIIT
NHSOGTVIGVHLIDGNAPEIIIGVGICIKLNAKISDFY"

BASE COUNT 309 a 318 c 353 g 310 t
Query Match 68.5%; Score 17.8; DB 3; Length 1290;
Best Local Similarity 90.5%; Pred. No. 3e+02; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGCTTTGCGAGATCGTACC 21
Db 996 CGCTTTGCGAGATCGTACC 976

RESULT 9
AF358993/c 1290 bp DNA linear INV 05-JUL-2001

LOCUS Trypanosoma cruzi strain TUI8 cl2 trypanothione reductase (TR)
DEFINITION gene, partial cds.

ACCESSION AF358993.1 GI:14600023

VERSION AF358993.1

KEYWORDS

SOURCE

ORGANISM

Trypanosoma cruzi.
Trypanosoma cruzi.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 1290)
Machado, C.A. and Ayala, F.J.
Nucleotide sequences provide evidence of genetic exchange among
distantly related lineages of Trypanosoma cruzi
Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)

JOURNAL

21310003
MEDLINE
11416213
2 (bases 1 to 1290)
Machado, C.A. and Ayala, F.J.
Direct Submission
Submitted (12-MAR-2001) Ecology and Evolutionary Biology,
University of California, Irvine, 321 Steinhaus Hall, Irvine, CA
92697-2525, USA

FEATURES
Source
Location/Qualifiers
1..1290
/organism="Trypanosoma cruzi"
/strain="TUI8 cl2"
/db_xref="taxon:5693"
<1..>1290
/gene="TR"
<1..>1290
/product="trypanothione reductase"
<1..>1290
/gene="TR"
/codon_start=1
/product="trypanothione reductase"
/protein_id="AAK71211.1"
/db_xref="GI:14600024"

gene
CDS
ORIGIN

309 a 318 c 353 g 310 t
BASE COUNT
ORIGIN
Query Match 68.5%; Score 17.8; DB 3; Length 1290;
Best Local Similarity 90.5%; Pred. No. 3e+02; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 CGCTTTGCGAGATCGTACC 21

||||||| ||| |||||
Db 996 CGCTTTGCGAGATCGTACC 976

RESULT 10
AF358994/c

LOCUS Trypanosoma cruzi strain CBB cl3 trypanothione reductase (TR) gene,
DEFINITION partial cds.

ACCESSION AF358994.1 GI:14600025

VERSION AF358994.1

KEYWORDS

SOURCE

ORGANISM

Trypanosoma cruzi.
Trypanosoma cruzi.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 1290)
Machado, C.A. and Ayala, F.J.
Nucleotide sequences provide evidence of genetic exchange among
distantly related lineages of Trypanosoma cruzi
Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001).

JOURNAL
21310003
MEDLINE
11416213
2 (bases 1 to 1290)
Machado, C.A. and Ayala, F.J.
Direct Submission
Submitted (12-MAR-2001) Ecology and Evolutionary Biology,
University of California, Irvine, 321 Steinhaus Hall, Irvine, CA
92697-2525, USA

FEATURES
Source
Location/Qualifiers
1..1290
/organism="Trypanosoma cruzi"
/strain="CBB cl3"
/db_xref="taxon:5693"
<1..>1290
/gene="TR"
<1..>1290
/product="trypanothione reductase"
<1..>1290
/gene="TR"
/codon_start=1
/product="trypanothione reductase"
/protein_id="AAK71212.1"
/db_xref="GI:14600026"

gene
CDS
ORIGIN

309 a 318 c 353 g 310 t
BASE COUNT
ORIGIN
Query Match 68.5%; Score 17.8; DB 3; Length 1290;
Best Local Similarity 90.5%; Pred. No. 3e+02; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 CGCTTTGCGAGATCGTACC 21

Db 996 CGCTTTGCGAGATCGTACC 976

RESULT 11
AF358995/c

LOCUS Trypanosoma cruzi strain MSC2 trypanothione reductase (TR) gene,
DEFINITION partial cds.

ACCESSION AF358995.1 GI:14600027

VERSION AF358995.1

KEYWORDS

SOURCE
ORGANISM Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE
AUTHORS Machado, C.A. and Ayala, F.J.
TITLE Trypanosoma; Schizotrypanum.
1 (bases 1 to 1290)
Nucleotide sequences provide evidence of genetic exchange among
distantly related lineages of Trypanosoma cruzi
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)
MEDLINE 21310003
PUBMED 11416213
REFERENCE
AUTHORS Machado, C.A. and Ayala, F.J.
TITLE Trypanosoma; Schizotrypanum.
1 (bases 1 to 1290)
Nucleotide sequences provide evidence of genetic exchange among
distantly related lineages of Trypanosoma cruzi
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)
MEDLINE 21310003
PUBMED 11416213
REFERENCE
2 (bases 1 to 1290)

FEATURES
Source
1. 1290
/organism="Trypanosoma cruzi"
/strain="MSC2"
/db_xref="taxon:5693"
/gene="tr"
/gene="tr"
/gene="tr"
/product="trypanothione reductase"
/protein_id="AAK71215.1"
/db_xref="GI:14600028"
/translation="TLKKRVAIVDQVWGHGPPRSALGCTCVNVCVPRKLMVYGAQ
GSLSEKNVNVRESADPASAVKERELETENILASGSMHPMIPGIEHCISSEAFIL
PEPPRRLVWGGFTSVFAGIFNATKREDDQVTLVYGEMLNGFPHITLREELTOL
TANGIOILKRENDARKVELNADSKSVTESGKMDFDVMAAIGRSPRTDLOLONQ
VMKNGVOVDVSRVSRVSNITAIQDVNRVMLPVAINEAALVDVPEGTNPKRTH
TRYASVPSIPPIGTGGLIEVASRKYEVAVVLSFPLMHNISGSKYRTFAKITT
NMSDGVAVHLGDNAPETIIGVGICIKLNAKISDFY"
BASE COUNT 309 a 310 t 353 g
ORIGIN

Query Match 68.5%; Score 17.8; DB 3; Length 1290;
Best Local Similarity 90.5%; Pred No. 3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCTCTTGGCAGATCGGAC 21
|||||
Db 996 CGCTCTTGGCAGATCGGAC 976

RESULT 12
AF358997/c 1290 bp DNA linear INV 05-JUL-2001
LOCUS Trypanosoma cruzi strain S03 c15 haplotype 2 trypanothione
DEFINITION reducease (tr) gene, partial cds.
ACCESSION AF358997.1 GI:14600031
VERSION AF358997.1 GI:14600031
KEYWORDS
SOURCE Trypanosoma cruzi.
ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE
AUTHORS Machado, C.A. and Ayala, F.J.
TITLE Trypanosoma; Schizotrypanum.
1 (bases 1 to 1290)
Nucleotide sequences provide evidence of genetic exchange among
distantly related lineages of Trypanosoma cruzi
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)
MEDLINE 21310003
PUBMED 11416213
REFERENCE
2 (bases 1 to 1290)

AUTHORS Machado, C.A. and Ayala, F.J.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2001) Ecology and Evolutionary Biology,
University of California, Irvine, 321 Steinhaus Hall, Irvine, CA
92697-2525, USA
FEATURES
Source
1. 1290
/organism="Trypanosoma cruzi"
/strain="S03 c15"
/db_xref="taxon:5693"
/haplotype="2"
/gene="tr"
/gene="tr"
/gene="tr"
/product="trypanothione reductase"
/protein_id="AAK71215.1"
/db_xref="GI:14600032"
/translation="TLKKRVAIVDQVWGHGPPRSALGCTCVNVCVPRKLMVYGAQ
GSLSEKNVNVRESADPASAVKERELETENILASGSMHPMIPGIEHCISSEAFIL
PEPPRRLVWGGFTSVFAGIFNATKREDDQVTLVYGEMLNGFPHITLREELTOL
TANGIOILKRENDARKVELNADSKSVTESGKMDFDVMAAIGRSPRTDLOLONQ
VMKNGVOVDVSRVSRVSNITAIQDVNRVMLPVAINEAALVDVPEGTNPKRTH
TRYASVPSIPPIGTGGLIEVASRKYEVAVVLSFPLMHNISGSKYRTFAKITT
NMSDGVAVHLGDNAPETIIGVGICIKLNAKISDFY"
BASE COUNT 309 a 310 t 353 g
ORIGIN

Query Match 68.5%; Score 17.8; DB 3; Length 1290;
Best Local Similarity 90.5%; Pred No. 3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCTCTTGGCAGATCGGAC 21
|||||
Db 996 CGCTCTTGGCAGATCGGAC 976

RESULT 13
AF358999/c 1290 bp DNA linear INV 05-JUL-2001
LOCUS Trypanosoma cruzi strain PSC-O haplotype 2 trypanothione reductase
DEFINITION (tr) gene, partial cds.
ACCESSION AF358999.1 GI:14600035
VERSION AF358999.1 GI:14600035
KEYWORDS
SOURCE Trypanosoma cruzi.
ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE
AUTHORS Machado, C.A. and Ayala, F.J.
TITLE Trypanosoma; Schizotrypanum.
1 (bases 1 to 1290)
Nucleotide sequences provide evidence of genetic exchange among
distantly related lineages of Trypanosoma cruzi
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)
MEDLINE 21310003
PUBMED 11416213
REFERENCE
2 (bases 1 to 1290)
Nucleotide sequences provide evidence of genetic exchange among
distantly related lineages of Trypanosoma cruzi
JOURNAL Submitted (12-MAR-2001) Ecology and Evolutionary Biology,
University of California, Irvine, 321 Steinhaus Hall, Irvine, CA
92697-2525, USA
FEATURES
Source
1. 1290
/organism="Trypanosoma cruzi"
/strain="PSC-O"
/db_xref="taxon:5693"
/haplotype="2"

gene
/gene="tr"
/gene="tr"
/product="trypanothione reductase"
/gene="tr"
/codon_start=1
/product="trypanothione reductase"
/protein_id="AAK71217.1"
/db_xref="GI:14600036"
BASE COUNT 309 a 318 c 353 g 310 t

ORIGIN
Query Match 68.5%; Score 17.8; DB 3; Length 1290;
Best Local Similarity 90.5%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCTTTGGCAGATCGGTACC 21
|||||
Db 996 CGCTTTGGCAGATCGGTACC 976

RESULT 14
AF359001/c 1290 bp DNA linear INV 05-JUL-2001
LOCUS
DEFINITION Trypanosoma cruzi strain CL F11F5 haplotype 2 trypanothione reductase (TR) gene, partial cds.
ACCESSION AF359001
VERSION AF359001.1 GI:14600039
KEYWORDS
SOURCE Trypanosoma cruzi.
ORGANISM Trypanosoma cruzi.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
1 (bases 1 to 1290)
Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of Trypanosoma cruzi
Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)
JOURNAL 2 (bases 1 to 1290)
PUBMED 11416213
MACHADO, C.A. and Ayala, F.J.
TITLE Direct Submission
Submitted (12-MAR-2001) Ecology and Evolutionary Biology,
University of California, Irvine, 321 Steinhaus Hall, Irvine, CA
92697-2525, USA
FEATURES
source
1.1290
Location/Qualifiers
/organism="Trypanosoma cruzi"
/strain="CL F11F5; CL Brener"
/db_xref="taxon:5693"
/haplotype="2"
/gene="tr"
/gene="tr"
/gene="tr"
/product="trypanothione reductase"
/codon_start=1
/product="trypanothione reductase"
/protein_id="AAK71219.1"
/db_xref="GI:14600040"

/translation="TLYKRRVAVIDVOMVHGPFPFSAAGTCVNVGCVPKKIMVTGAO
YMEHLRESAGFMEFDRITTLRAEMKKLIAVDEAVLNINKSYEEMFRTGEEFLGW
GSLSEKNVNVNRESADPASAVKERELETENILASGMPHMPNIPGIEHCISSEAFYL
PEPPRRVLTGGGFIISVEFAGIFNAYKPKDQVTLCTYGEMLIRGFDHRLRELLKOL
TANGIOILITKENPAKVELNADSGSKVTSKGMDFVMAAIGSPRTKDLQONAG
VMKNGVQVDEYSRTNVSNIYALIGDVNRVMTVAINEAALVDVFGTNPRTDH
TRVAVAFSIPIGTCGLIEEVAASKRYEVAVYLSFPLMHNINISGSXYKTFVAKIIT
NHSDDTVLGVLHGNAPRIIOGVGICLKLNKISDFY"
BASE COUNT 309 a 318 c 353 g 310 t

ORIGIN
Query Match 68.5%; Score 17.8; DB 3; Length 1290;
Best Local Similarity 90.5%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCTTTGGCAGATCGGTACC 21
|||||
Db 996 CGCTTTGGCAGATCGGTACC 976

RESULT 15
AF359003/c 1290 bp DNA linear INV 05-JUL-2001
LOCUS
DEFINITION Trypanosoma cruzi strain Tulahuen c12 haplotype 2 trypanothione reductase (TR) gene, partial cds.
ACCESSION AF359003
VERSION AF359003.1 GI:14600043
KEYWORDS
SOURCE Trypanosoma cruzi.
ORGANISM Trypanosoma cruzi.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
1 (bases 1 to 1290)
Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of Trypanosoma cruzi
Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)
JOURNAL 2 (bases 1 to 1290)
PUBMED 11416213
MACHADO, C.A. and Ayala, F.J.
TITLE Direct Submission
Submitted (12-MAR-2001) Ecology and Evolutionary Biology,
University of California, Irvine, 321 Steinhaus Hall, Irvine, CA
92697-2525, USA
FEATURES
source
1.1290
Location/Qualifiers
/organism="Trypanosoma cruzi"
/strain="Tulahuen c12"
/db_xref="taxon:5693"
/haplotype="2"
/gene="tr"
/gene="tr"
/gene="tr"
/product="trypanothione reductase"
/codon_start=1
/product="trypanothione reductase"
/protein_id="AAK71221.1"
/db_xref="GI:14600044"
/translation="TLYKRRVAVIDVOMVHGPFPFSAAGTCVNVGCVPKKIMVTGAO
YMEHLRESAGFMEFDRITTLRAEMKKLIAVDEAVLNINKSYEEMFRTGEEFLGW
GSLSEKNVNVNRESADPASAVKERELETENILASGMPHMPNIPGIEHCISSEAFYL
PEPPRRVLTGGGFIISVEFAGIFNAYKPKDQVTLCTYGEMLIRGFDHRLRELLKOL
TANGIOILITKENPAKVELNADSGSKVTSKGMDFVMAAIGSPRTKDLQONAG
VMKNGVQVDEYSRTNVSNIYALIGDVNRVMTVAINEAALVDVFGTNPRTDH
TRVAVAFSIPIGTCGLIEEVAASKRYEVAVYLSFPLMHNINISGSXYKTFVAKIIT
NHSDDTVLGVLHGNAPRIIOGVGICLKLNKISDFY"
BASE COUNT 309 a 318 c 353 g 310 t

Fri Jul 11 15:29:38 2003

us-10-087-631b-8.rge

Page 12

Query Match 68.5%; Score 17.8; DB 3; Length 1290;
Best Local Similarity 90.5%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGTCTTCGCGAGATCGGTACC 21
|||||
Db 996 CGTCTTCGCGAGATCGGTACC 976

Search completed: July 10, 2003, 19:20:30
Job time : 232.294 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 54.8629 Seconds
(without alignments)
1067.241 Million cell updates/sec

Title: US-10-087-631B-8

Perfect score: 26

Sequence: 1 cgccttcgcagatcgtactcaat 26

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

N.GeneSeq_101002:*

1: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT:*

2: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT:*

3: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT:*

4: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT:*

5: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT:*

6: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT:*

7: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT:*

8: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT:*

9: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT:*

10: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT:*

11: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT:*

12: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT:*

13: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT:*

14: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT:*

15: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT:*

16: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT:*

17: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT:*

18: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT:*

19: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT:*

20: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT:*

21: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT:*

22: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA2001A.DAT:*

23: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT:*

24: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.6	67.7	768	23	ABL30111
2	17.6	67.7	1215	21	AAC48208
3	17.6	67.7	2112	23	ABL12431
4	17.6	67.7	2378	23	ABL28118
5	17.6	67.7	2902	23	ABL30110
6	17.6	67.7	3522	13	AAQ26928
7	17.6	67.7	3704	23	ABL30114
8	17.6	67.7	4233	23	ABL12430
9	17.6	67.7	8173	23	ABL13658

c	10	17	65.4	7407	23	ABL10262
c	11	17	65.4	34980	22	AAH41225
c	12	17	65.4	3077	21	AAA46570
c	13	16.6	63.8	3092	14	AAQ48934
c	14	16.6	63.8	3092	14	ABK49453
c	15	16.6	63.8	3126	16	AAQ90112
c	16	16.6	63.8	3202	21	AACT8077
c	17	16.6	63.8	3213	24	AAI39859
c	18	16.6	63.8	3213	24	ABL62632
c	19	16.6	63.8	3253	14	AAQ48935
c	20	16.6	63.8	3253	14	ABK49494
c	21	16.6	63.8	3324	22	AAAD12765
c	22	16.6	63.8	3448	22	AAH72618
c	23	16.6	63.8	3477	22	AAI26761
c	24	16.6	63.8	10409	19	AAV42558
c	25	16.4	63.1	480	21	AAC53967
c	26	16.4	63.1	617	24	ABQ66243
c	27	16.4	63.1	689	21	AAC35194
c	28	16.4	63.1	720	21	AAFI2664
c	29	16.4	63.1	1794	23	AA55491
c	30	16.4	63.1	2502	18	AAH88991
c	31	16.4	63.1	11470	23	ABL21296
c	32	16.4	63.1	11770	19	AAV52305
c	33	16.2	62.3	1026	20	AAK61466
c	34	16.2	62.3	4798	22	AAK65321
c	35	16.2	62.3	443	24	ABO56215
c	36	16.1	61.5	447	21	AAFI0043
c	37	16.1	61.5	614	24	ABO66261
c	38	16.1	61.5	705	18	AAV25142
c	39	16.1	61.5	705	23	AA535748
c	40	16.1	61.5	734	21	AAC81059
c	41	16.1	61.5	1168	21	AAC37451
c	42	16.1	61.5	1194	21	AAC48789
c	43	16.1	61.5	1195	21	AAC38594
c	44	16.1	61.5	1608	21	AAC42319
c	45	16.1	61.5	1794	24	ABK35748

ALIGNMENTS

RESULT 1

ABL30111 standard: DNA: 768 BP.

ABL30111:

AC ABL30111:

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide seq ID NO 41806.

XX Drosophila: developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX OS

XX PN

XX MO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US092331.

XX PF

XX 23-MAR-2000; 2000US-191637P.

XX PR

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

Drosophila melanog
Pyrococcus abyssi
cDNA of different
hosp-2p1. Homo sa
DNA encoding OSF-2
TCI gene. Homo sa
Human cancer assoc
Human allerge-asso
Colon adenocarcino
hosp-2os. Homo sa
DNA encoding OSF-2
Human OSF-2 (perio
Human cervical can
Human breast cance
Mouse dectin-2 gen
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Aspergillus oryzae
Streptococcus pneu
Streptococcus pneu
Drosophila melanog
Streptococcus pneu
DNA encoding a hum
Human immune/haema
Human ovarian anti
Fusarium venenatum
Arabidopsis thalia
H. pylori cytoplas
Helicobacter pylori
Human secreted pro
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
cDNA sequence #139

PT Interactions
XX Claim 1; SEQ ID NO 41806; 21bp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
CC sequences (AB57737-AB72072).
CC (AB57737-AB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
SO Sequence 768 BP; 173 A; 214 C; 218 G; 163 T; 0 other;
Query Match 67.7%; Score 17.6; DB 23; Length 768;
Best Local Similarity 83.3%; Pred. No. 49;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 2 GTCTTTCGACATCGTACTCA 25
Db 600 GTCTTCGCGGTCTACTCA 623
RESULT 2
AAC48208/c ID AAC48208 standard; DNA: 1215 BP.
AC AAC48208;
XX
DT 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 56648.
DE
XX Hybridisation assay; genetic mapping; gene expression control;
KW Protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD
XX 06-SEP-2000.
XX
PF 25-FEB-2000; 200EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 99US-0134218.

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 16-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

```

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149126.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155659.
PR 24-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

```

```

PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 67.7%; Score 17.6; DB 21; Length 1215;
 Best Local Similarity 83.3%; Pred. No. 52;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 3 TCTTCGACGACGCTACCTCAAT 26
    ||||| ||||| ||||| |||||
DB 773 TCTTATCATGATGGACCTCAAT 750

```

RESULT 3
 ID ABL12431 standard; cDNA; 2112 BP.

XX ABL12431;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31775.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB68328.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions

XX Claim 1: SEQ ID NO 31775; 21np + sequence listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 capable of detecting 1000 or more genes from Drosophila. The invention is
 useful in developmental biology and in elucidating cell signalling and
 cell-cell interactions in higher eukaryotes for the development of
 insecticides, therapeutics and pharmaceutical drugs. The invention
 discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 sequences (ABL01840-ABL16175) and the encoded proteins
 (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2112 BP; 542 A; 555 C; 583 G; 432 T; 0 other;

Query Match 67.7%; Score 17.6; DB 23; Length 2112;
 Best Local Similarity 83.3%; Pred. No. 55;

26-MAR-2002 16:45:00

XX 1.3

PD 18-JUN-1992.

XX 05-JUN-1990; 90AU-0056291.
 PF Best Local Similarity 83.3%; Score 17.6; DB 13; Length 3522;
 XX 05-JUN-1990; 90AU-0056291.
 XX (MYCO) MYCOGEN CORP.
 PA WPI: 1992-259364/32.
 DR P-PSDB: ABR25825.
 XX
 PT Controlling insect pests of lepidoptera family - includes
 PT contacting insects with suitable amt. of Bacillus thuringiensis
 PS Claim 13; Page 29; 49pp; English.
 XX
 CC The DNA sequence encodes a Bacillus thuringiensis endotoxin which
 CC may be used in an expression system to transform a variety of
 CC microbial hosts, e.g. Pseudomonas, Azobacter, Erwinia, Serratia,
 CC Agrobacterium, Streptomyces species, etc. Expression of the toxic
 CC gene results in the intracellular prodn. and maintenance of the
 CC peptide toxin. With suitable hosts, the microbes can be applied to
 CC the sites of lepidopteran insects, e.g. to the insects themselves,
 CC to the rhizosphere, phylloplane or to a body of water, where they
 CC will proliferate and be ingested by the insects. The result is the
 CC control of unwanted insects. Alternatively, the microbe hosting the
 CC toxin gene can be treated under conditions that prolong the
 CC activity of the toxin produced in the cell. The treated cell can
 CC be applied to the environment of target pests. The resulting prod.
 CC retains the toxicity of the B.t. toxin. See also AA026929-31.
 CC
 XX Sequence 3522 BP; 1171 A; 579 C; 766 G; 1006 T; 0 other;
 SQ
 Query Match 67.7%; Score 17.6; DB 13; Length 3522;
 Best Local Similarity 83.3%; Pred. No. 59;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 3 TCTTTCGACATCGTACTCAAT 26
 ||||| ||||| ||||| |||||
 DB 252 TCTTTCGACATCGTACTCAAT 229
 RESULT 7
 ID ABL30114/C
 AC ABL30114 standard; DNA; 3704 BP.
 XX
 AC ABL30114;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 41815.
 XX
 KM Drosophila: developmental biology; cell signalling; insecticide;
 KM pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI: 2001-656860/75.
 DR
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -
 XX
 PS Claim 1; SEQ ID NO 41815; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB101840-AB16175) and the encoded proteins
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pt_sequences.
 CC
 XX Sequence 3704 BP; 782 A; 963 C; 961 G; 998 T; 0 other;
 SQ
 Query Match 67.7%; Score 17.6; DB 23; Length 3704;
 Best Local Similarity 83.3%; Pred. No. 59;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 2 GTCCTTCGACATCGTACTCA 25
 ||||| ||||| ||||| |||||
 DB 3419 GTCCTTCGACATCGTACTCA 3396
 RESULT 8
 ID ABL12430/C
 AC ABL12430 standard; CDNA; 4233 BP.
 XX
 AC ABL12430;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31772.
 XX
 KM Drosophila: developmental biology; cell signalling; insecticide;
 KM pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI: 2001-656860/75.
 DR
 XX P-PSDB: ABB68327.
 PT
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Claim 1; SEQ ID NO 31772; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB101840-AB16175) and the encoded proteins
 CC The sequence data for this patent did not form part of the printed

CC Specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4233 BP; 1150 A; 979 C; 934 G; 1170 T; 0 other;

Query Match

Best Local Similarity 67.7%; Score 17.6; DB 23; Length 4233;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

XX 2 GCTTTTCGAGATCGGTACTCTCAA 25
 DB 3009 GCTTTTCGAGATCGGTACTCTCAA 2986

RESULT 9

ABL13658/c

ID ABL13658 standard; cDNA; 8173 BP.

XX ABL13658;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 35456.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

FD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP. NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB69555.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions

XX Claim 1; SEQ ID NO 35456; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 8173 BP; 2452 A; 1765 C; 1688 G; 2268 T; 0 other;

Query Match

Best Local Similarity 67.7%; Score 17.6; DB 23; Length 8173;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

XX 2 GCTTTTCGAGATCGGTACTCTCAA 25
 DB 204 GCTTTTCGAGATCGGTACTCTCAA 181

RESULT 10

ABL10262/c

ID ABL10262 standard; cDNA; 7407 BP.

XX ABL10262;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 25268.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP. NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB66159.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions

XX Claim 1; SEQ ID NO 25268; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 7407 BP; 1810 A; 1780 C; 1770 G; 2047 T; 0 other;

Query Match

Best Local Similarity 65.4%; Score 17; DB 23; Length 7407;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

XX 1 GCTTTTCGAGATCGGTACTCTCAA 25
 DB 313 GCTTTTCGAGATCGGTACTCTCAA 289

RESULT 11

AAH41225/c

ID AAH41225 standard; DNA; 349980 BP.

XX AAH41225;

DT 29-OCT-2001 (first entry)

DE Pyrococcus abyssi genomic fragment #4.

XX Hyperthermophilic archaeon; hyperthermophilic protein; ds.

OS Pyrococcus abyssi.
 XX Key Location/Qualifiers
 FH misc-feature 1.49980
 FT /tag= a
 FT /note= "This sequence overlaps with the 3' end of
 FT AAA41224"
 FT misc-feature 300001..349980
 FT /tag= b
 FT /note= "This sequence overlaps with the 5' end of
 FT AAA41226"
 PN FR2792651-A1.
 PD 27-OCT-2000.
 XX 21-APR-1999; 99FR-0005034.
 PF 21-APR-1999; 99FR-0005034.
 XX 21-APR-1999; 99FR-0005034.
 XX (CNRS) CNRS CENT NAT RECH SCI
 PA (IFREMER) IFREMER INST FR RECH EXPL MER.
 XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 PI Querellou J, Weissenbach J, Saurin W, Hellig R;
 XX WPI: 2001-126236/14.
 DR New nucleotide sequences isolated from Pyrococcus abyssi encode
 PT proteins useful in industry -
 XX Claim 1: Page 429-524; 1657pp; French.
 XX The present invention relates to the genomic sequence of Pyrococcus
 CC abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
 CC hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
 CC vents. The present invention is a fragment of the genomic sequence of P.
 CC abyssi. The 5' end of this sequence overlaps with the 3' end of AAA41224
 CC and the 3' end of this sequence overlaps with the 5' end of AAA41226. The
 CC proteins of the present invention have various potential industrial uses,
 CC since the proteins are stable at very high temperatures, some up to 110
 CC degrees centigrade.
 CC Note: This patent is in the same patent family as WO200065062, which
 CC contains additional sequences as shown in AAB99132-AAB99143,
 CC AAH75903-AAH75920 and AAG66436.
 XX Sequence 349980 BP; 99421 A; 76616 C; 77444 G; 96499 T; 0 other;
 SQ
 Query Match 65.4%; Score 17; DB 22; Length 349980;
 Best Local Similarity 80.0%; Pred. No. 2e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 1 CGCTTTGCGAGATCGTACTCTCAA 25
 DB 78251 CGTCATCAGCGATCGGCACCTCAA 78227
 RESULT 12
 AAA46670
 ID AAA46670 standard; CDNA; 3077 BP.
 XX AAA46670;
 AC
 XX 25-SEP-2000 (first entry)
 DT
 XX CDNA of differentially expressed human gene osf-2.
 DE
 XX Differentially expressed human gene; cardiac disease; kidney disease;
 KW inflammatory disease; I-8u; prostacyclin-stimulating factor; tsf-2;
 KW tissue specific mRNA; insulin-like growth factor binding protein 6;
 KW OSF-1; gas-1; YMP; BRG2; pre-B cell stimulating factor homologue; SDF1a;
 KW peripheral benzodiazepine receptor; annexin II cellular ligand; p11;
 KW congenital heart failure; dilated congestive cardiomyopathy;

KW hypertrophic cardiomyopathy; restrictive cardiomyopathy;
 KW mitral valve disease; aortic valve disease; tricuspid valve disease;
 KW angina pectoris; myocardial infarction; cardiac arrhythmia;
 KW pulmonary hypertension; arterial hypertension; renovascular hypertension;
 KW arteriosclerosis; atherosclerosis; cardiac tumour; ss.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH CDS 28.2367
 FT /tag= a
 FT
 PN WO200035473-A2.
 PD 22-JUN-2000.
 XX 15-DEC-1999; 99WO-US29941.
 PF 15-DEC-1999; 99WO-US29941.
 XX 18-DEC-1998; 98US-0113008.
 XX (SCIO-) SCIOS INC.
 PA Stanton LW, White RT, Damm DL, Lewicki JA, Joly A, Schreiner GF;
 PI WPI: 2000-451904/39.
 DR P-PSDB: AAY93596.
 XX Preventing, diagnosing and treating cardiac, kidney and inflammatory
 PT disorders using cardiac genes that are differentially expressed in
 PT disease states such as cardiac arrhythmia and arteriosclerosis -
 XX Disclosure: Fig 7C; 170pp; English.
 XX AAA46668-79 represent differentially expressed human genes, associated
 CC with disease states and disorders. The specification describes
 CC methods preventing, diagnosing and treating cardiac, kidney and
 CC inflammatory diseases associated with inappropriate expression of
 CC differentially expressed cardiac, kidney and inflammatory genes
 CC (e.g. AAA46668-79). These genes include I-8u, prostacyclin-stimulating
 CC factor, tsf-2, tissue specific mRNA, insulin-like growth factor
 CC binding protein 6, OSF-1, gas-1, YMP, BRG2, pre-B cell stimulating
 CC factor homologue (SDF1a), peripheral benzodiazepine receptor, and
 CC cellular ligand of annexin II (p11), respectively. These diseases
 CC include congenital heart failure, dilated congestive cardiomyopathy,
 CC hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve
 CC disease, aortic valve disease, tricuspid valve disease, angina pectoris,
 CC myocardial infarction, cardiac arrhythmia, pulmonary hypertension,
 CC arterial hypertension, renovascular hypertension, arteriosclerosis,
 CC atherosclerosis and/or cardiac tumours.
 XX Sequence 3077 BP; 1032 A; 571 C; 632 G; 842 T; 0 other;
 SQ
 Query Match 63.8%; Score 16.6; DB 21; Length 3077;
 Best Local Similarity 82.6%; Pred. No. 1.8e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 4 CTTTGGCAGATCGTACTCTCAAT 26
 DB 715 CTTACCAAAATTGGTACTCTCAAT 737
 RESULT 13
 AAQ48934
 ID AAQ48934 standard; CDNA to mRNA; 3092 BP.
 XX AAQ48934;
 AC
 XX 15-APR-1994 (first entry)
 DT
 XX hOSF-2p1.
 DE
 XX Bone-related protein; bone; diagnosis; disease; growth factor;
 KW cell adhesion; guiding; induction; metabolic bone disease; ss.

XX OS Homo sapiens (placenta).
 XX XX
 XX Key Location/Qualifiers
 XX CDS 38..2377
 FT sig_peptide /tag= a
 FT sig_peptide /product= OSF-2
 FT sig_peptide 38..101
 FT mat_peptide /tag= b
 FT mat_peptide 102..2374
 FT /tag= c
 XX
 XX EP562508-A.
 XX
 XX PD 29-SEP-1993.
 XX
 XX PF 22-MAR-1993; 93EP-0104650.
 XX
 XX PR 27-MAR-1992; 92JP-0071501.
 XX
 XX PA (FARH) HOECHST JAPAN LTD.
 XX
 XX PI Anann E, Kikuno R, Otawara-Hamamoto Y, Takeshita S;
 XX Tezuka K;
 XX WPI: 1993-304910/39.
 XX DR P-PSDB; AAR41868.
 XX
 XX PT Mammalian OSF-2 protein - obtd. from bone tissue, useful for
 XX diagnosis and treatment of metabolic bone disease
 XX
 XX PS Claim 6; Page 14-18; 58pp; English.
 XX
 XX CC OSF-2 plays an important role in the formation of bone, by acting as
 XX a growth factor or adhesion or "guiding" protein to attract cells
 XX to the site of bone induction. In a pharmaceutical compn. it can
 XX be applied in metabolic bone diseases.
 XX
 XX SQ Sequence 3092 BP; 1035 A; 573 C; 638 G; 846 T; 0 other;
 XX
 XX Query Match 63.8%; Score 16.6; DB 14; Length 3092;
 XX Best Local Similarity 82.6%; Pred. No. 1.8e+02;
 XX Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 XX
 XX OY 4 CTTTGCAGATCGTACTCAAT 26
 XX DB 725 CTTACACAAATTGTGACTCTCAAT 747
 XX
 XX RESULT 14
 XX ID ABR49493 standard; DNA; 3092 BP.
 XX AC ABR49493;
 XX
 XX DT 15-JUL-2002 (first entry)
 XX
 XX DE DNA encoding OSF-2 protein associated protein #3.
 XX
 XX KW Heart failure; OSF-2 protein; gene therapy; transgenic animal;
 XX human; gene; ds.
 XX
 XX OS Homo sapiens.
 XX
 XX XX
 XX FT Key Location/Qualifiers
 XX CDS 38..2377
 FT /tag= a
 FT /product= "OSF-2 associated protein #3"
 XX
 XX PD 14-MAR-2002.
 XX
 XX

PF 07-SEP-2001; 2001WO-JP07787.
 XX
 XX XX
 XX PR 08-SEP-2000; 2000JP-0273644.
 XX
 XX PA (SUNR) SUNTORI LTD.
 XX (SUNT-) SUNTORI BIOMEDICAL RES LTD.
 XX
 XX PI Kawashima K, Katsuragi N, Sugimura K, Furuya M, Morishita R;
 XX WPI: 2002-371858/40.
 XX DR P-PSDB; AAU79825.
 XX
 XX PT Remedies for heart failure obtained by screening drugs capable of
 XX inhibiting expression of OSF-2 gene or production or function of the
 XX protein -
 XX
 XX PS Claim 5; Page 61-66; 80pp; Japanese.
 XX
 XX CC The invention describes preventives or remedies for heart failure
 XX comprising a substance inhibiting the expression of an OSF (undefined)-2
 XX protein or function of the protein or function of a target molecule of
 XX the OSF-2 protein. The remedies are for heart failure, and are also
 XX applicable for gene therapy. Diagnosis can also be achieved by monitoring
 XX gene expression and protein production. Transgenic animals are useful for
 XX studying the degree of disease deterioration and onset risk forecast to
 XX provide treatment and help adjust lifestyle. This sequence encodes an
 XX OSF-2 associated protein.
 XX
 XX SQ Sequence 3092 BP; 1035 A; 573 C; 638 G; 846 T; 0 other;
 XX
 XX Query Match 63.8%; Score 16.6; DB 24; Length 3092;
 XX Best Local Similarity 82.6%; Pred. No. 1.8e+02;
 XX Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 XX
 XX OY 4 CTTTGCAGATCGTACTCAAT 26
 XX DB 725 CTTACACAAATTGTGACTCTCAAT 747
 XX
 XX RESULT 15
 XX ID AA090112 standard; cDNA; 3126 BP.
 XX AC AA090112;
 XX
 XX DT 05-NOV-1995 (first entry)
 XX
 XX DE TCI gene.
 XX
 XX KW Tumour marker; invasive; metastatic; cancer; ss.
 XX
 XX OS Homo sapiens.
 XX
 XX FT Key Location/Qualifiers
 XX CDS 43..2376
 FT /tag= a
 FT
 XX
 XX PN WO9511923-A.
 XX
 XX PD 04-MAY-1995.
 XX
 XX PF 31-OCT-1994; 94WO-US12502.
 XX
 XX PR 29-OCT-1993; 93US-0146488.
 XX
 XX PA (DAND) DANA FARMER CANCER INST INC.
 XX
 XX PI Bao S, Chen LB, Liu Y;
 XX WPI: 1995-178826/23.
 XX DR P-PSDB; AAR74302.
 XX
 XX PT New tumour marker TCI, corresp. DNA and monoclonal antibody - for

PT detecting, preventing and treating tumours, esp. in breast, colon
PT and gastrointestinal tract cancer.

PS Disclosure; Flg 4; 84pp; English.

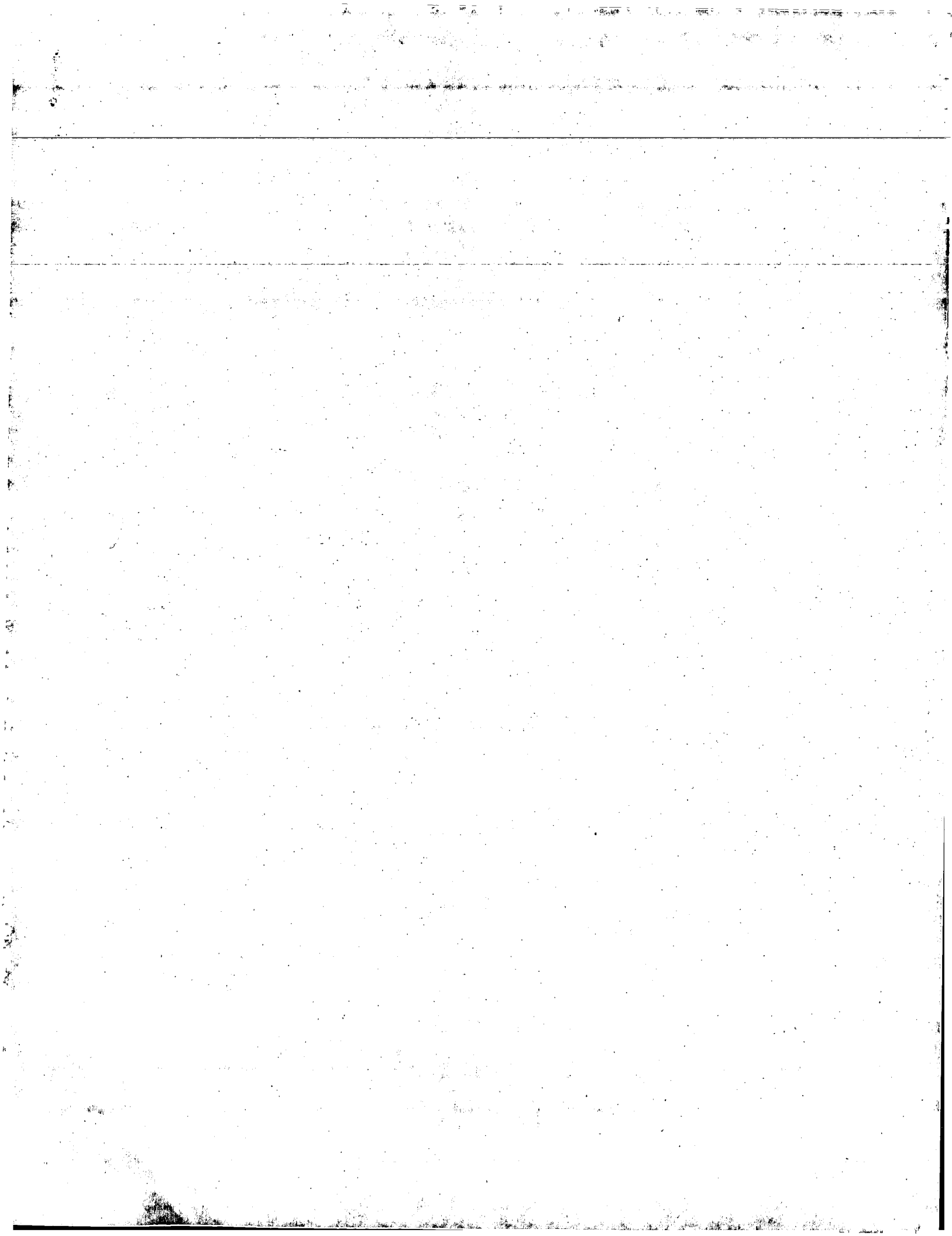
CC The sequence is that of the TCI gene which encodes the TCI tumour
CC marker protein. The gene and its product may be used to detect
CC tumours in blood, urine or sputum. Inhibitors of TCI are used to
CC treat late stage cancers and for preventing tumour cell metastasis.
CC See also MAQ90113-25.

SQ Sequence 3126 BP; 1035 A; 611 C; 656 G; 824 T; 0 other;

Query Match:	63.8%	Score	16.6	DB	16.	Length	3126.
Best Local Similarity	82.6%	Pred	No.	1.8e+02.			
Matches	19.	Conservative	0.	Mismatches	4.	Indels	0.
						Gaps	0.

QY	4	CTTTGCAGATCGGTACCTCAAT	26
Db	748	CTTACACAATTTGGTACCTCAAT	770

Search completed: July 10, 2003, 19:52:45
Job time : 57.8629 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 11.627 Seconds

(Without alignments)
685,785 Million cell updates/sec

Title: US-10-087-631b-8

Perfect score: 26

Sequence: 1 cgccttcgcagcgcgtactcaat 26

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2.6/ptodata/1/ina/5A.COMB.seq:*

2: /cgn2.6/ptodata/1/ina/5B.COMB.seq:*

3: /cgn2.6/ptodata/1/ina/6A.COMB.seq:*

4: /cgn2.6/ptodata/1/ina/6B.COMB.seq:*

5: /cgn2.6/ptodata/1/ina/PCUTS.COMB.seq:*

6: /cgn2.6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	65.4	495	4	US-09-221-017B-1098
2	16.6	63.8	636	2	US-08-477-396A-1
3	16.6	63.8	3092	1	US-08-426-627-3
4	16.6	63.8	3126	2	US-08-477-396A-3
5	16.6	63.8	3253	1	US-08-426-627-5
6	16.6	63.8	10409	3	US-08-772-440-33
7	16.4	63.1	2502	2	US-08-844-086-1
8	16.4	63.1	2502	2	US-09-018-211-1
9	16.4	63.1	11770	4	US-08-961-527-172
10	16	61.5	2373	4	US-08-975-762-45
11	16	61.5	2373	4	US-09-295-028-45
12	16	61.5	2373	4	US-09-106-582-45
13	16	61.5	2574	4	US-09-253-829-28
14	16	61.5	3522	1	US-08-040-751-4
15	16	61.5	3522	1	US-08-291-368-1
16	16	61.5	3522	2	US-08-962-190-1
17	16	61.5	3522	5	PCR-US95-10310-1
18	16	61.5	3522	6	5164180-3
19	16	61.5	4003	1	US-08-257-999-1
20	16	61.5	7195	4	US-08-478-507-6
21	16	61.5	7195	4	US-09-128-275A-6
22	16	61.5	7195	4	US-09-553-427-6
23	15.6	60.0	4635	4	US-09-221-017B-187
24	15.4	59.2	1200	4	US-09-502-653-7
25	15.4	59.2	1875	2	US-08-683-743-3
26	15.4	59.2	99500	4	US-09-798-096-10
27	15.2	58.5	44	1	US-08-344-695-5

28	15.2	58.5	580	4	US-09-328-111-465	Sequence 465, App
29	15.2	58.5	1224	1	US-07-718-274A-5	Sequence 5, Appl
30	15.2	58.5	1224	1	US-08-149-106-5	Sequence 5, Appl
31	15.2	58.5	1224	1	US-08-298-021-5	Sequence 5, Appl
32	15.2	58.5	1586	1	US-07-841-646-18	Sequence 18, Appl
33	15.2	58.5	1586	1	US-08-147-023-18	Sequence 18, Appl
34	15.2	58.5	1586	1	US-08-447-570-18	Sequence 18, Appl
35	15.2	58.5	1586	2	US-08-449-700-18	Sequence 18, Appl
36	15.2	58.5	1586	2	US-08-449-699A-18	Sequence 18, Appl
37	15.2	58.5	1751	1	US-08-377-292-4	Sequence 4, Appl
38	15.2	58.5	1788	1	US-07-841-646-6	Sequence 6, Appl
39	15.2	58.5	1788	1	US-07-901-703-18	Sequence 6, Appl
40	15.2	58.5	1788	1	US-08-147-023-6	Sequence 6, Appl
41	15.2	58.5	1788	1	US-08-447-570-6	Sequence 6, Appl
42	15.2	58.5	1788	2	US-08-449-700-6	Sequence 6, Appl
43	15.2	58.5	1788	2	US-08-449-699A-6	Sequence 6, Appl
44	15.2	58.5	1788	5	PCR-US93-05446-18	Sequence 18, Appl
45	15.2	58.5	1954	1	US-08-050-132A-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-221-017B-1098
Sequence 1098, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: ROSS, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq. for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
APPLICATION NUMBER: PP2911
APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 1098:
SEQUENCE CHARACTERISTICS:
LENGTH: 495 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1...495
US-09-221-017b-1098

Query Match 65.4%; Score 17; DB 4; Length 495;
Best Local Similarity 80.0%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGCTCTTCGACATCGTACTCTCAA 25
DB 228 CGACTGTACGATCGTACTCTCAA 252

RESULT 2

US-08-477-396A-1
Sequence 1, Application US/08477396A
Patent No. 5872235

GENERAL INFORMATION:
APPLICANT: Chen, Ian Bo
APPLICANT: Bao, Shideng
APPLICANT: Liu, Yuan
TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: Weingarten, Schungtin, Gagnebin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477.396A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/146,488
FILING DATE: 29-OCT-1993
APPLICATION NUMBER: US 08/448,388
FILING DATE: 28-MAY-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/12502
FILING DATE: 31-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Heine, Holliday C.
REGISTRATION NUMBER: 34,346
REFERENCE/DOCKET NUMBER: DFCI-333BX
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-2290
TELEFAX: (617) 451-0313
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..636

US-08-477-396A-1

Query Match 63.8%; Score 16.6; DB 2; Length 636;
Best Local Similarity 82.6%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CTTTCGACATCGTACTCTCAAT 26
DB 58 CTTACACAAATTTGTACTCTCAAT 80

RESULT 3

US-08-426-627-3
Sequence 3, Application US/08426627
Patent No. 575664

GENERAL INFORMATION:
APPLICANT: Amann, Egon
APPLICANT: Otawara-Hamamoto, Yoko
APPLICANT: Kikuno, Reiko
APPLICANT: Takeshita, Sunao
APPLICANT: Tezuka, Kenichi
TITLE OF INVENTION: NO. 575664e1 Protein with Bone Formation
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSER: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/426,627
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/036,841
FILING DATE: 25-MAR-1993
APPLICATION NUMBER: JP 4-71501
FILING DATE: 27-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hammond, Alan W.
REGISTRATION NUMBER: 35,178
REFERENCE/DOCKET NUMBER: 02481-1285-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4400
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3092 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Placenta
FEATURE:
NAME/KEY: CDS
LOCATION: join(38..2375)
FEATURE:
NAME/KEY: mat-peptide
LOCATION: join(101..2375)
US-08-426-627-3

Query Match 63.8%; Score 16.6; DB 1; Length 3092;
Best Local Similarity 82.6%; Pred. No. 29;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CTTGGCAGATCGTACTCAAT 26

DB 725 CTTACACAATGTGTACTCAAT 747

RESULT 4

US-08-477-396A-3

; Sequence 3, Application US/08477396A

; Patent No. 5872235

; GENERAL INFORMATION:

; APPLICANT: Chen, Ian Bo

; APPLICANT: Bao, Shideng

; APPLICANT: Liu, Yuan

; TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes

; STREET: Ten Post Office Square

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/477,396A

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/146,488

; FILING DATE: 29-OCT-1993

; APPLICATION NUMBER: US 08/448,388

; FILING DATE: 28-MAY-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/12502

; FILING DATE: 31-OCT-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Heine, Holliday C.

; REGISTRATION NUMBER: 34,346

; REFERENCE/DOCKET NUMBER: DFCI-333BX

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-2290

; TELEFAX: (617) 451-0313

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3126 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 43..2376

; US-08-477-396A-3

Query Match 63.8%; Score 16.6; DB 2; Length 3126;

Best Local Similarity 82.6%; Pred. No. 29;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CTTGGCAGATCGTACTCAAT 26

DB 748 CTTACACAATGTGTACTCAAT 770

RESULT 5

US-08-426-627-5

; Sequence 5, Application US/08426627

; Patent No. 5756664

; GENERAL INFORMATION:

; APPLICANT: Amano, Egon

; APPLICANT: Otawara-Hamamoto, Yoko

; APPLICANT: Kikuno, Reiko

; APPLICANT: Takeshita, Sunao

; APPLICANT: Tezuka, Kenichi

; TITLE OF INVENTION: No. 5756664e1 Protein with Bone Formation

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finegan, Henderson, Farabow, Garrett &

; ADDRESS: Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/426,627

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/036,841

; FILING DATE: 25-MAR-1993

; APPLICATION NUMBER: JP 4-71501

; FILING DATE: 27-MAR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Hammond, Alan W.

; REGISTRATION NUMBER: 35,178

; REFERENCE/DOCKET NUMBER: 02481-1285-00000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-408-4000

; TELEFAX: 202-408-4400

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3253 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; CELL TYPE: osteosarcoma

; FEATURE:

; NAME/KEY: CDS

; LOCATION: join(32..2540)

; FEATURE:

; NAME/KEY: mat.peptide

; LOCATION: join(97..2540)

; US-08-426-627-5

Query Match 63.8%; Score 16.6; DB 1; Length 3253;

Best Local Similarity 82.6%; Pred. No. 29;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CTTGGCAGATCGTACTCAAT 26

DB 719 CTTACACAATGTGTACTCAAT 741

RESULT 6

US-08-772-440-33

; Sequence 33, Application US/08772440

; Patent No. 6046158

; GENERAL INFORMATION:

APPLICANT: Arizumi, Kiyoshi
APPLICANT: Takashima, Akira
TITLE OF INVENTION: UNIQUE DENDRITIC CELL-ASSOCIATED C-TYPE
TITLE OF INVENTION: LECTINS, DECTIN-1 AND DECTIN-2; COMPOSITIONS AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/772,440
FILING DATE: CONCURRENTLY HERewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTXD:493
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 10409 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: modified_base
LOCATION: 6510
OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "D = A or G or T"
FEATURE:
NAME/KEY: modified_base
LOCATION: 3406..6470
OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "K = G or T"
FEATURE:
NAME/KEY: modified_base
LOCATION: 3564..7896
OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "M = A or C"
FEATURE:
NAME/KEY: modified_base
LOCATION: 3497..3607
OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "N = A or C or G or T"
FEATURE:
NAME/KEY: modified_base
LOCATION: 3479..6422
OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "R = A or G"
FEATURE:
NAME/KEY: modified_base
LOCATION: 3405..6871
OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "S = C or G"
FEATURE:
NAME/KEY: modified_base
LOCATION: 3457..9958
OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "W = A or T"
FEATURE:
NAME/KEY: modified_base
LOCATION: 3595..9959

OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "Y = C or T"
US-08-772-440-33
Query Match
Best Local Similarity 79.2%; Score 16.6; DB 3; Length 10409;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 3 TCTTGGCAGATCGGTACTCAAT 26
DB 2852 TCTTGGCAGATCGGTACTCAAT 2875
RESULT 7
US-08-844-086-1
Sequence 1, Application US/08844086
Patent No. 586390
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 586390el Compounds
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Smithline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fastseq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,086
FILING DATE: 18-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-844-086-1
Query Match
Best Local Similarity 76.9%; Score 16.4; DB 2; Length 2502;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CGCTTTGGCAGATCGGTACTCAAT 26
DB 590 CGCTTTGGCAGATCGGTACTCAAT 615
RESULT 8
US-09-018-211-1
Sequence 1, Application US/09018211
Patent No. 6048716
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 6048716el Compounds

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Smithline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,211
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/844,086
FILING DATE: 18-APR-1997
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-018-211-1

Query Match 63.1%; Score 16.4; DB 3; Length 2502;
Best Local Similarity 76.9%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CGCTTTGCGAGATCGTACTCAAT 26
DB 590 CGGCTTACGACAGCGCTTCTCAAT 615

RESULT 9
US-08-961-527-172
Sequence 172: Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MS-DOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brooks, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 172:
SEQUENCE CHARACTERISTICS:
LENGTH: 11770 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-172

Query Match 63.1%; Score 16.4; DB 4; Length 11770;
Best Local Similarity 76.9%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CGCTTTGCGAGATCGTACTCAAT 26
DB 7818 CGGCTTACGACAGCGCTTCTCAAT 7843

RESULT 10
US-08-975-762-45
Sequence 45: Application US/08975762
Patent No. 6207169
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,762
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 2373 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-975-762-45

Query Match 61.5%; Score 16; DB 4; Length 2373;
Best Local Similarity 79.2%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 CGCTTTGCGAGATCGTACTCA 24

DB 910 CATCTTTCGCGATAGTACTCA 933

RESULT 11

US-09-295-028-45
Sequence 45, Application US/09295028
Patent No. 6277381

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
FILE REFERENCE: 210121.439C4

CURRENT APPLICATION NUMBER: US/09/295,028
CURRENT FILING DATE: 1999-04-20
NUMBER OF SEQ ID NOS: 85

SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 45

LENGTH: 2373

TYPE: DNA
ORGANISM: Ehrlichia sp.

US-09-295-028-45

Query Match
Best Local Similarity 79.2%; Score 16; DB 4; Length 2373;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGCTTTTCGCGATCGTACTCA 24
DB 910 CATCTTTCGCGATAGTACTCA 933

RESULT 12

US-09-106-582-45
Sequence 45, Application US/09106582
Patent No. 6306402

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:

THERAPY

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582

FILING DATE: 29-JUN-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Mark, David J.

REGISTRATION NUMBER: 31,392

TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900

TELEFAX: 206-682-6031

INFORMATION FOR SEQ ID NO: 45:

SEQUENCE CHARACTERISTICS:
LENGTH: 2373 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-106-582-45

Query Match
Best Local Similarity 79.2%; Score 16; DB 4; Length 2373;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGCTTTTCGCGATCGTACTCA 24
DB 910 CATCTTTCGCGATAGTACTCA 933

RESULT 13

US-09-255-829-28/c
Sequence 28, Application US/09255829
Patent No. 6461617

GENERAL INFORMATION:

APPLICANT: Shone, Clifford Charles
APPLICANT: Quinn, Conrad Padraig
APPLICANT: Foster, Keith Alan

TITLE OF INVENTION: Recombinant Toxin Fragments
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/255,829

FILING DATE: 23-FEB-1999

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB97/02273

FILING DATE: 22-AUG-1997

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/782,893

FILING DATE: 27-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: ESMOND, ROBERT W.

REGISTRATION NUMBER: 32,893

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2540

TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:
LENGTH: 2574 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..2574

US-09-255-829-28

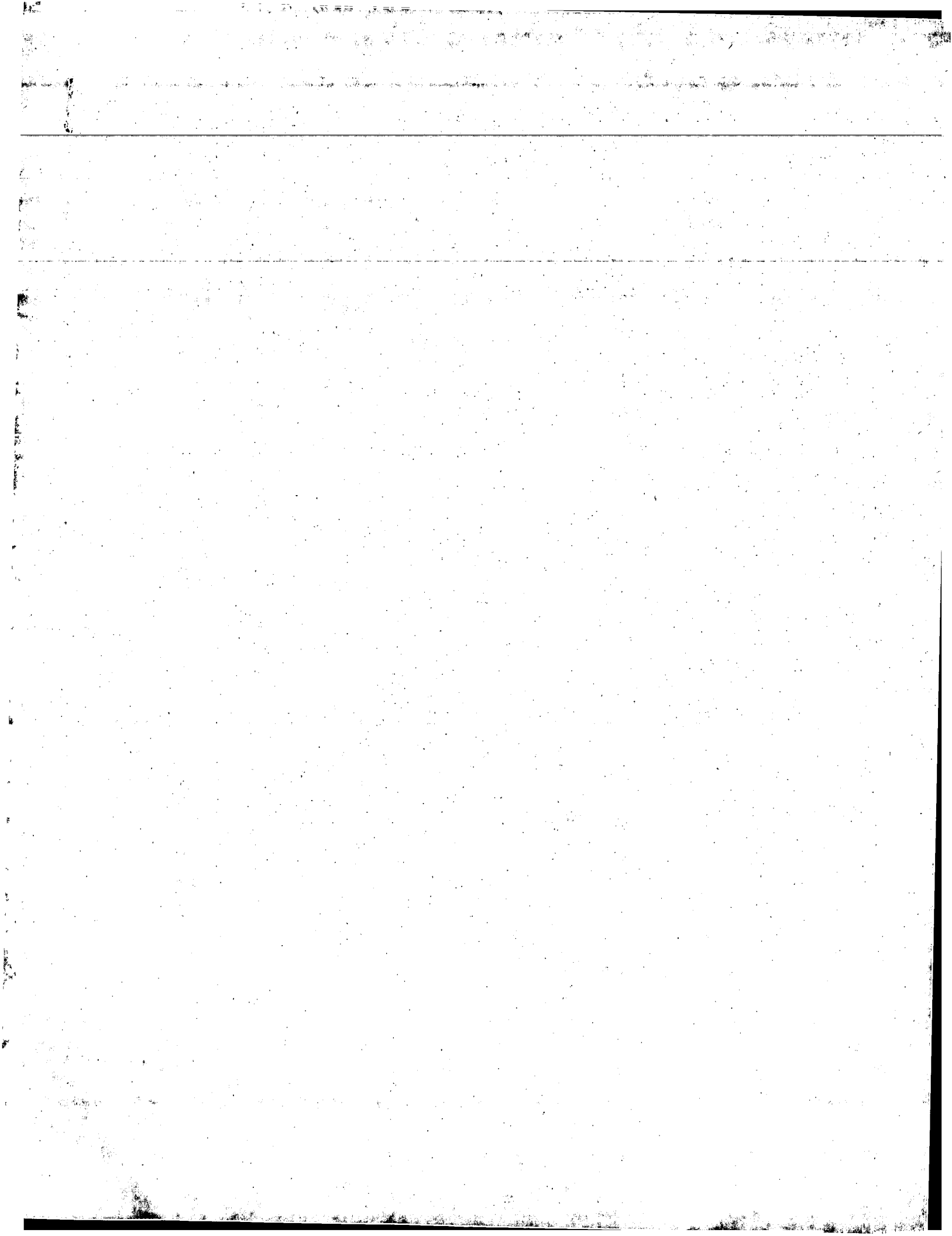
Query Match
Best Local Similarity 61.5%; Score 16; DB 4; Length 2574;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GCTTTTCGCGATCGTACTCA 25
DB 1570 GTTTTGTAGACCGGAGCTCA 1547

RESULT 14

US-08-040-751-4/c
Sequence 4, Application US/08040751
Patent No. 5407825
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M
TITLE OF INVENTION: No. 5407825el Bacillus thuringiensis isolates
TITLE OF INVENTION: active against Lepidopteran Pests and Genes Encoding No. 54078
TITLE OF INVENTION: Lepidopteran-active Toxins
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DAVID R. SALIWANCHIK
STREET: 2421 N.W. 41st STREET, SUITE A-1
CITY: GAINESVILLE
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/040.751
FILING DATE: 19930329
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SALIWANCHIK, DAVID R.
REFERENCE/DOCKET NUMBER: MA39.C1.D3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3522 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
STRAIN: aizawai
INDIVIDUAL ISOLATE: PS81A2
IMMEDIATE SOURCE:
LIBRARY: Lambdagem - 11 (tm) Library of August Sick
CLONE: 81A2
US-08-040-751-4
Query Match 61.5%; Score 16; DB 1; Length 3522;
Best Local Similarity 79.2%; Pred. No. 58;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 3 TCTTTCGCGATCGGTACTCAAT 26
||||| 11 1111 11 111111
DB 252 TCTTGGCGCATCACTCAAT 229
RESULT 15
US-08-291-368-1/c
Sequence 1, Application US/08291368
Patent No. 5686069
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M.
TITLE OF INVENTION: No. 5686069el Bacillus thuringiensis Isolates
TITLE OF INVENTION: Active Against Lepidopteran Pests
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: SalIWanchik & SalIWanchik
STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/291.368
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/597,607
FILING DATE: 15-OCT-90
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: SalIWanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA50.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904)375-8100
TELEFAX: (904)372-5800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3522 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
STRAIN: aizawai
INDIVIDUAL ISOLATE: PS81A2
IMMEDIATE SOURCE:
LIBRARY: Lambdagem - 11 (tm) Library of August Sick
CLONE: 81A2
US-08-291-368-1
Query Match 61.5%; Score 16; DB 1; Length 3522;
Best Local Similarity 79.2%; Pred. No. 58;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 3 TCTTTCGCGATCGGTACTCAAT 26
||||| 11 1111 11 111111
DB 252 TCTTGGCGCATCACTCAAT 229
Search completed: July 10, 2003, 20:27:41
Job time : 13.627 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: July 10, 2003, 22:53:16 ; Search time 71.164 Seconds
(without alignments)
576.891 Million cell updates/sec

Title: US-10-087-631B-8

Perfect score: 26
Sequence: 1 cgccttcgcagatcgctacccaat 26

Scoring table: IDENTITY NUC
Gapop 10.0 ; Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCF_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/PCFUS_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	26	US-10-087-631B-8	Sequence 8, Appl1
2	24.4	99.8	241	US-10-087-631B-12	Sequence 12, Appl1
3	24.4	99.8	943	US-10-087-631B-11	Sequence 11, Appl1
4	17	65.4	25	US-10-088-263B-53770	Sequence 53770, A
5	16.6	63.8	507	US-09-796-692-4795	Sequence 4795, Ap
6	16.6	63.8	507	US-10-040-862-4795	Sequence 4795, Ap
7	16.6	63.8	2190	US-10-217-371-5	Sequence 4, Appl1
8	16.6	63.8	2250	US-10-217-371-13	Sequence 13, Appl1
9	16.6	63.8	2253	US-10-217-371-3	Sequence 3, Appl1
10	16.6	63.8	2274	US-10-217-371-9	Sequence 9, Appl1
11	16.6	63.8	2313	US-10-217-371-11	Sequence 11, Appl1
12	16.6	63.8	2337	US-10-217-371-7	Sequence 7, Appl1
13	16.6	63.8	2508	US-10-217-371-1	Sequence 1, Appl1
14	16.6	63.8	2937	US-10-176-847-45	Sequence 45, Appl1
15	16.6	63.8	3077	US-10-171-311-119	Sequence 179, App
16	16.6	63.8	3202	US-09-925-301-171	Sequence 471, App
17	16.6	63.8	3213	US-10-171-311-177	Sequence 177, App
18	16.6	63.8	3213	US-10-177-293-341	Sequence 341, App
19	16.6	63.8	3681	US-10-198-846-9861	Sequence 9861, Ap

c	20	16.6	63.8	397658	10	US-09-813-320-3	Sequence 3, Appl1
	21	16.6	63.8	9025608	9	US-10-156-761-1	Sequence 1, Appl1
	22	16.4	63.1	617	10	US-09-770-149-820	Sequence 820, App
	23	16.4	63.1	1794	10	US-09-815-242-9128	Sequence 9128, Ap
	24	16.4	63.1	5289	9	US-10-128-714-179	Sequence 179, App
	25	16.4	63.1	5289	9	US-10-128-714-5179	Sequence 5179, Ap
	26	16.4	63.1	25	9	US-10-215-112-6258	Sequence 6258, Ap
	27	16	61.5	168	10	US-09-878-574-13775	Sequence 13775, A
	28	16	61.5	490	9	US-09-918-995-33372	Sequence 33372, A
	29	16	61.5	614	10	US-09-770-149-838	Sequence 838, App
	30	16	61.5	705	10	US-09-815-242-7385	Sequence 7385, Ap
	31	16	61.5	945	9	US-10-084-546-5	Sequence 5, Appl1
	32	16	61.5	1410	9	US-09-938-842A-818	Sequence 818, App
	33	16	61.5	1794	9	US-09-822-846-139	Sequence 139, App
	34	16	61.5	2184	10	US-09-801-368-295	Sequence 295, App
	35	16	61.5	2373	10	US-09-159-469-45	Sequence 45, Appl1
	36	16	61.5	2373	10	US-09-798-042-45	Sequence 45, Appl1
	37	16	61.5	3120	9	US-10-128-714-261	Sequence 261, App
	38	16	61.5	3120	9	US-10-128-714-5261	Sequence 5261, App
	39	16	61.5	3969	9	US-10-177-293-386	Sequence 386, App
	40	16	61.5	4235	9	US-10-198-846-10905	Sequence 10905, A
	41	16	61.5	4416	9	US-10-128-714-395	Sequence 395, App
	42	16	61.5	4417	9	US-10-128-714-5395	Sequence 5395, App
	43	16	61.5	7195	9	US-09-851-410-6	Sequence 6, Appl1
	44	15.8	60.8	135	10	US-09-878-574-7595	Sequence 7595, App
	45	15.8	60.8	333	10	US-09-294-093B-124	Sequence 124, App

ALIGNMENTS

RESULT 1
US-10-087-631B-8
Sequence 8, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
TITLE OF INVENTION: CONTROL
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087, 631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8
LENGTH: 26
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: 5780pc primer parallel
US-10-087-631B-8

Query Match 100.0%; Score 26; DB 9; Length 26;
Best Local Similarity 100.0%; Pred.No. 0.0014;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCTTTCGACATCGCTACCAAT 26
DB 1 CGCTTTCGACATCGCTACCAAT 26

RESULT 2
US-10-087-631B-12
Sequence 12, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
TITLE OF INVENTION: CONTROL
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087, 631B
CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 12
LENGTH: 241
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Amplicon derived from QS (pc)
US-10-087-631B-12

Query Match
Best Local Similarity 93.8%; Score 24.4; DB 9; Length 241;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGCTTCGCGAGATCGTACTCCTCAAT 26
1 CGCTTCGCGAGATCGTACTCCTCAAT 26

RESULT 3.
US-10-087-631B-11
Sequence 11, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 11
LENGTH: 943
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: QS (pc) HCV being parallel-
US-10-087-631B-11

Query Match
Best Local Similarity 93.8%; Score 24.4; DB 9; Length 943;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGCTTCGCGAGATCGTACTCCTCAAT 26
57 CGCTTCGCGAGATCGTACTCCTCAAT 82

RESULT 4
US-10-098-263B-53770/c
Sequence 53770, Application US/10098263B
Publication No. US20030104410A1
GENERAL INFORMATION:
APPLICANT: Mittman, Michael
TITLE OF INVENTION: Human Microarray
FILE REFERENCE: 3118.1
CURRENT APPLICATION NUMBER: US/10/098,263B
CURRENT FILING DATE: 2003-01-08
PRIOR APPLICATION NUMBER: 60/276,759
PRIOR FILING DATE: 2001-03-16
NUMBER OF SEQ ID NOS: 131066
SOFTWARE: Microarray Probe-Sequence Listing Generator V 1.1
SEQ ID NO 53770
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapiens
US-10-098-263B-53770

Query Match
Best Local Similarity 65.4%; Score 17; DB 9; Length 25;
Matches 80.0%; Pred. No. 46; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 GCTTCGCGAGATCGTACTCCTCAAT 26
25 GTACTCGCGAGATCGTACTCCTCAAT 1

RESULT 5
US-09-796-692-4795
Sequence 4795, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
FILE REFERENCE: 2077.001200
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/223,378
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4795
LENGTH: 507
TYPE: DNA
ORGANISM: Homo sapiens
US-09-796-692-4795

Query Match
Best Local Similarity 63.8%; Score 16.6; DB 9; Length 507;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CTTTCGCGAGATCGTACTCCTCAAT 26
438 CTTACACAAATTTGTTACTCCTCAAT 460

RESULT 6
US-10-040-862-4795
Sequence 4795, Application US/10040862
Publication No. US20030078396A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
APPLICANT: Retter, Marc
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther
TITLE OF INVENTION: Hematological Malignancies

```
FILE REFERENCE: 014058-013520US
CURRENT APPLICATION NUMBER: US/10/040,862
CURRENT FILING DATE: 2001-11-06
PRIOR APPLICATION NUMBER: US 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: US 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: US 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: US 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: US 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: US 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: US 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: US 60/223,378
PRIOR FILING DATE: 2000-08-07
PRIOR APPLICATION NUMBER: US 09/796,692
PRIOR FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 10467
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4795
LENGTH: 507
TYPE: DNA
ORGANISM: Homo sapiens
US-10-040-862-4795

Query Match          63.8%; Score 16.6; DB 9; Length 507;
Best Local Similarity 82.6%; Pred. No. 94;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CTTGCGAGATCGGTACTCTCAAT 26
DB 438 CTTACACAATTTGTACTCTCAAT 460

RESULT 7
US-10-217-371-5
; Sequence 5, Application US/10217371
; Publication No. US20030073137A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Ian Bo
; APPLICANT: Dai, Meiru
; APPLICANT: Sasaki, Hidefumi
; APPLICANT: Auciailt, Daniel
; TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS
; FILE REFERENCE: 00530-099001
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/10/217,371
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US 60/312,123
; PRIOR FILING DATE: 2001-08-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2190
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2190)
US-10-217-371-5

Query Match          63.8%; Score 16.6; DB 9; Length 2190;
```

```
Best Local Similarity 82.6%; Pred. No. 11e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CTTGCGAGATCGGTACTCTCAAT 26
DB 625 CTTACACAATTTGTACTCTCAAT 647

RESULT 8
US-10-217-371-13
; Sequence 13, Application US/10217371
; Publication No. US20030073137A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Ian Bo
; APPLICANT: Dai, Meiru
; APPLICANT: Sasaki, Hidefumi
; APPLICANT: Auciailt, Daniel
; TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS
; FILE REFERENCE: 00530-099001
; CURRENT APPLICATION NUMBER: US/10/217,371
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US 60/312,123
; PRIOR FILING DATE: 2001-08-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 2250
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2250)
US-10-217-371-13

Query Match          63.8%; Score 16.6; DB 9; Length 2250;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CTTGCGAGATCGGTACTCTCAAT 26
DB 625 CTTACACAATTTGTACTCTCAAT 647

RESULT 9
US-10-217-371-3
; Sequence 3, Application US/10217371
; Publication No. US20030073137A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Ian Bo
; APPLICANT: Dai, Meiru
; APPLICANT: Sasaki, Hidefumi
; APPLICANT: Auciailt, Daniel
; TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS
; FILE REFERENCE: 00530-099001
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/10/217,371
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US 60/312,123
; PRIOR FILING DATE: 2001-08-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2253
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2253)
US-10-217-371-3

Query Match          63.8%; Score 16.6; DB 9; Length 2253;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

OY 4 CTTTCGAGATGGTACTCAAT 26
DB 688 CTTACACAATTTGTACTCTCAAT 710

RESULT 10

US-10-217-371-9
Sequence 9, Application US/10217371
Publication No. US20030073137A1
GENERAL INFORMATION:
APPLICANT: Chen, Ian Bo
APPLICANT: Dai, Meiru
APPLICANT: Sasaki, Hidefumi
TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS
FILE REFERENCE: 00530-099001
CURRENT APPLICATION NUMBER: US/10/217,371
PRIOR FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: US 60/312,123
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 2274
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(2274)
US-10-217-371-9

Query Match
Best Local Similarity 63.8%; Score 16.6; DB 9; Length 2274;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 CTTTCGAGATGGTACTCTCAAT 26
DB 625 CTTACACAATTTGTACTCTCAAT 647

RESULT 11

US-10-217-371-11
Sequence 11, Application US/10217371
Publication No. US20030073137A1
GENERAL INFORMATION:
APPLICANT: Chen, Ian Bo
APPLICANT: Dai, Meiru
APPLICANT: Sasaki, Hidefumi
TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS
FILE REFERENCE: 00530-099001
CURRENT APPLICATION NUMBER: US/10/217,371
PRIOR FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: US 60/312,123
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 2313
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(2313)
US-10-217-371-11

Query Match
Best Local Similarity 63.8%; Score 16.6; DB 9; Length 2313;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 CTTTCGAGATGGTACTCTCAAT 26
DB 688 CTTACACAATTTGTACTCTCAAT 710

RESULT 12

US-10-217-371-7
Sequence 7, Application US/10217371
Publication No. US20030073137A1
GENERAL INFORMATION:
APPLICANT: Chen, Ian Bo
APPLICANT: Dai, Meiru
APPLICANT: Sasaki, Hidefumi
TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS
FILE REFERENCE: 00530-099001
CURRENT APPLICATION NUMBER: US/10/217,371
PRIOR FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: US 60/312,123
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 2337
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(2337)
US-10-217-371-7

Query Match
Best Local Similarity 63.8%; Score 16.6; DB 9; Length 2337;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 CTTTCGAGATGGTACTCTCAAT 26
DB 688 CTTACACAATTTGTACTCTCAAT 710

RESULT 13

US-10-217-371-1
Sequence 1, Application US/10217371
Publication No. US20030073137A1
GENERAL INFORMATION:
APPLICANT: Chen, Ian Bo
APPLICANT: Dai, Meiru
APPLICANT: Sasaki, Hidefumi
TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS
FILE REFERENCE: 00530-099001
CURRENT APPLICATION NUMBER: US/10/217,371
PRIOR FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: US 60/312,123
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2508
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(2508)
US-10-217-371-1

Query Match
Best Local Similarity 63.8%; Score 16.6; DB 9; Length 2508;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 CTTTCGAGATGGTACTCTCAAT 26
DB 688 CTTACACAATTTGTACTCTCAAT 710

RESULT 14

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:13:45 ; Search time 467.474 Seconds
(without alignments)
900.761 Million cell updates/sec

Title: US-10-087-631B-8

Sequence: 1 cgtcttcgcagatcgtactcaat 26

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmuv:*
5: em_estcov:*
6: em_estipl:*
7: em_estiro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.6	75.4	961	17	AO900643 HS_3171.A
2	19.2	73.8	185	10	BB600244 BB600244
3	19.2	73.8	698	13	BI924556 EST544445
4	18.6	71.5	460	9	AI351632 q05h12.x
5	18.6	71.5	1029	12	BF098973 601750475
6	18.2	70.0	182	17	CNS02NMJ CNS02NMJ

Result No.	Score	Query Match	Length	DB ID	Description
7	18.2	70.0	234	10	BB593576 BB593576
8	18.2	70.0	242	10	BB600933 BB600933
9	18.2	70.0	243	10	BB569672 BB569672
10	18.2	70.0	290	10	BB592646 BB592646
11	18.2	69.2	312	10	BB252299 BB252299
12	18.2	69.2	411	17	AA087346 HS_2191.B
13	18.2	69.2	435	12	BF588219 FMI_39.B0
14	18.2	69.2	701	17	AA055220 mqxb0009D
15	17.6	67.7	150	12	BF757321 MRO-C7045
16	17.6	67.7	285	13	BI172474 BI172474
17	17.6	67.7	335	10	AV524463 AV524463
18	17.6	67.7	381	9	AA538976 AA538976
19	17.6	67.7	386	14	BM870563 mgns010XH
20	17.6	67.7	421	17	AA147865 SP_0014.A
21	17.6	67.7	454	13	BI164958 BI164958
22	17.6	67.7	498	13	BI171702 BI171702
23	17.6	67.7	512	17	BI177119 BI177119
24	17.6	67.7	512	17	CNS07J0J CNS07J0J
25	17.6	67.7	527	13	BI227641 BI227641
26	17.6	67.7	547	13	BI170103 BI170103
27	17.6	67.7	557	13	BI362843 BI362843
28	17.6	67.7	578	9	AA263759 AA263759
29	17.6	67.7	578	10	AV517896 AV517896
30	17.6	67.7	581	13	BI213740 BI213740
31	17.6	67.7	589	13	BI168983 BI168983
32	17.6	67.7	590	14	BI16252 BI16252
33	17.6	67.7	594	9	AT515767 AT515767
34	17.6	67.7	594	9	AI999483 AI999483
35	17.6	67.7	599	13	BI169019 BI169019
36	17.6	67.7	605	13	BI171671 BI171671
37	17.6	67.7	611	13	BI356268 BI356268
38	17.6	67.7	618	9	AA820697 AA820697
39	17.6	67.7	624	13	BI485494 BI485494
40	17.6	67.7	624	13	BI230819 BI230819
41	17.6	67.7	642	13	BI169762 BI169762
42	17.6	67.7	659	13	BI238807 BI238807
43	17.6	67.7	659	17	BI175201 BI175201
44	17.6	67.7	659	17	CNS0718R CNS0718R
45	17.6	67.7	662	17	BB655017 BB655017

ALIGNMENTS

RESULT 1
LOCUS AO900643 961 bp DNA linear GSS 10-NOV-1999
DEFINITION HS_3171.A2.A07.TTC CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone plate-3171 COL-14 Row-A, DNA sequence.
ACCESSION AO900643
VERSION AO900643.1 GI:6356833
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 961)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL MEDLINE
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: <http://www.htsc.washington.edu>

FEATURES	Source
Plate: 3171	row: A column: 14
Seq primer: T7	
Class: BAC ends	
High quality sequence stop: 961	
location/qualifiers	
1. 961	

BASE COUNT ORIGIN					
146 a	415 c	143 g	257 t		

Query Match	75.48;	Score 19.6;	DB 17;	Length 961;
Best Local Similarity	84.68;	Pred. No. 1.5e+02;		
Matches	22;	Conservative	0;	Mismatches 4;
				Indels 0;
				Gaps 0

Dy 1 CGTCTTCGCAGATCGGTACCTCAAT 26
||||| ||||| ||||
Db 711 CGTCTGTCGGGTTCCGTACCACAAT 736

RESULT 2	BB600244/c	185 bp	mrna	linear	EST 01-DEC-2006
LOCUS	BB600244				
DEFINITION	BB600244	RIKEN full-length enriched, 12 days embryo			spinal ganglion
ACCESSION	Mus musculus cDNA clone D13004T24 5'				mrna sequence.
VERSION	BB600244.1				
KEYWORDS	GI:11508845				
EST.					

ORGANISM	REFERENCE	AUTHORS
<i>Mus musculus</i>		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Mus (bases 1 to 185)		
Alizawa, K.; Akhita, S.; Akimura, T.; Arai, A.; Arthaus, T.; Carratelli, F.		

REFERENCE
AUTHORS

1 (bases 1 to 185)
Aizawa, K., Akchira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, T.,
Hanagaki, T., Hayatsu, N., Hiraoka, T., Hirozane, T., Hodyoyama, Y.,
Imotoani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Kon,
H., Kusakabe, M., Matsuyama, T., Miyazaki, M., Nakamura, M., Nish, K.
Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C.,
Sakai, K., Saeki, D., Sato, K., Shibata, K., Shibata, Y., Shinagawa, A.,
Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka,
T., Toyota, T., Watabiki, A., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.
RIKEN MOUSE ESTS (Aizawa, K. et al. 2000)
Unpublished (2000)
CONTACT: Yoshinhide Hayashizaki

Laboratory for Genome Exploitation Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9210
Email: genome-res@sc.riken.go.jp,
url:http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasakawa, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotabilization and thermoactivation of the molable enzymes by trehalose and its application for the synthesis of full length cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)
Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES
SOURCE

```
Location/Qualifiers
1. 185
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D130004124"
/clone_lib="Riken full-length enriched, 12 days embryo
spinal ganglion"
/tissue-type="spinal ganglion"
/dev_stage="12 days embryo"
lab_host="m10days embryo"
```

BASE COUNT
BRIGIN

42 a 44 c 61 g 38 t

plunged with a primer [5' -GAGGAGAGACGGCGCCGCCACCTCGAGTCTTTTCTTTTTTTTTTTTAAVN 3'] , cDNA was prepared by using Trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper . Second strand cDNA was prepared with the primer adapter of sequence [5' -GAGGAGAGACTTCGTGCAGTTAATTAATTAAATCAATCCCCCCCCCCC 3'] . cDNA was cleaved with BamHI and XhoI . Vector : a modified pluscript KS(+) after bulk excision from lambda PLC I . "

Query Match	73.8%	Score 19.2	DB 10	Length 165
Best Local Similarity	87.5%	Pred. No. 1.7e+02		
Matches	21, conservative	0, Mismatches	3, Indels	0, Gaps

QY 3 TCTTTGCAGATCGGTACTCAAT 26
Db 48 TCTTTCACAAATCGGTACTTCAT 25

RESULT 3					
B1924556/c					
LOCUS	B1924556	698 bp	mRNA	linear	EST 18-OCT-2001
DEFINITION	EST534445 tomato flower, buds 0-3 mm				
ACCESSION	clone cTGA23N9 5' end, mRNA sequence.				
VERSION	B1924556				
KEYWORDS	B1924556.1 GI:16228164				
SOURCE	EST.				
ORGANISM	tomato.				
	Lycopersicon esculentum				

REFERENCES
AUTHORS

TITLE Matelli, G.B., Giovannoni, J.J., and Tanksley, S.D.
JOURNAL Generation of ESIs from tomato flower tissue, 0-3 mm buds (2001)
COMMENT Unpublished (2001)
Contact: CUGI

Features
 Source
 1. 698
 Location/Qualifiers
 Seq primer: T3.
 This clone is available through the Clemson University Genomics
 Institute
 Email: <http://www.genome.clemson.edu/orders/index.html>
 100 Jordan Hall, Clemson, SC 29634, USA
 Clemson University

```
Location/Qualifiers
1. 698
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CT0A23N9"
/clone_lib="tomato flower, buds 0-3 mm"
/tissue_type="flower"
```

```

/dev stage="0-3mm buds"
/Note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Cornell University; sequencing: The
Institute for Genomic Research; Flower buds and flowers
were taken from greenhouse plants (4-8 wks old, TA496).
They were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."
BASE COUNT      214 a      120 c      166 g      198 t
ORIGIN

Query Match      73.8%; Score 19.2; DB 13; Length 698;
Best Local Similarity 87.5%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 TCTTTCGACATCGTACTCAAT 26
      1 ||||| ||||| ||||| |||||
      133 TTTTTCGACATGGTACTCAAT 110

RESULT 4
LOCUS      A1351632      460 bp      mRNA      linear      EST 13-FEB-1999
DEFINITION      qf05h12.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
ACCESSION      A1351632
IMAGE:1940039 3', mRNA sequence.
VERSION      A1351632.1 GI:4088838
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 460)
AUTHORS      NCI-CCAP http://www.nci.nlm.nih.gov/ncicgap.
TITLES      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
      Email: cgaabs-remail.nih.gov
      This clone is available royalty-free through LNL; contact the
      IMAGE Consortium (info@image.llnl.gov) for further information.
      Insert Length: 1001 Std Error: 0.00
      Seq primer: -400p from Gldco
      High quality sequence stop: 437.
FEATURES
SOURCE      Location/Qualifiers
      1..460
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_image="IMAGE:1940039"
      /clone_lib="Soares_total_fetus_Nb2HF8_9w"
      /dev stage="8-9 weeks"
      /lab host="DH10B"
      /note="Vector: pT733-Pac (Pharmacia) with a modified
      polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
      was prepared from mRNA obtained from pooled 8-9 week
      (total) fetus material with a Not I - oligo(dT) primer [5'
      TGTACCAATCTGAAGAGGAGCGCCGCTTATTTTCTTTTCTTTT 3'].
      Double-stranded cDNA was ligated to Eco RI adaptors
      (Pharmacia), digested with Not I and cloned into the Not I
      and Eco RI sites of the modified pT733 vector. Library
      went through one round of normalization, and was
      constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      131 a      91 c      107 g      131 t
ORIGIN

Query Match      71.5%; Score 18.6; DB 9; Length 460;
Best Local Similarity 84.0%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 GTCTTTCGACATCGTACTCAAT 26
      ||||| ||||| ||||| |||||
      422 GTCTTTCGACATCGTACTCAAT 398

```

```

RESULT 5
LOCUS      BF098973/c      1029 bp      mRNA      linear      EST 19-OCT-2000
DEFINITION      601750475p1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3978234 5',
mRNA sequence.
ACCESSION      BF098973
VERSION      BF098973.1 GI:10881499
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 1029)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLES      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
      Email: cgaabs-remail.nih.gov
      Tissue Procurement: Gilbert Smith, Ph.D.
      cDNA Library Preparation: Life Technologies, Inc.
      cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LNL at:
      http://image.llnl.gov
      Plate: LLM9170 row: e column: 19
      High quality sequence stop: 636.
FEATURES
SOURCE      Location/Qualifiers
      1..1029
      /organism="Mus musculus"
      /strain="FVB/N"
      /db_xref="taxon:10090"
      /clone_image="IMAGE:3978234"
      /clone_lib="NCI_CGAP_Mam1"
      /tissue_type="tumor, biopsy sample"
      /dev stage="10 months, virgin"
      /lab host="DH10B"
      /note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI;
      Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
      Library constructed by Life Technologies. Investigator
      providing samples: Gilbert Smith, NIH"
BASE COUNT      257 a      295 c      300 g      177 t
ORIGIN

Query Match      71.5%; Score 18.6; DB 12; Length 1029;
Best Local Similarity 84.0%; Pred. No. 4.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 GTCTTTCGACATCGTACTCAAT 26
      ||||| ||||| ||||| |||||
      825 GTCTTTCGACATCGTACTCAAT 801

RESULT 6
LOCUS      CNS02NM/c      182 bp      DNA      linear      GSS 14-MAY-2000
DEFINITION      Tetraodon nigroviridis genome survey sequence T7 end of clone
151019 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION      CNS02NM
VERSION      AL205483
KEYWORDS      GSS: genome survey sequence.
SOURCE      Tetraodon nigroviridis.
ORGANISM      Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE      1 (bases 1 to 182)
AUTHORS      Roest-Crollius, R., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fitzames, C., Wincker, P., Brothier, P., Queller, F.,
Saurin, W. and Weissbach, J.
TITLES      Human gene number estimate provided by genome wide analysis using

```

JOURNAL Tetraodon nigroviridis DNA sequence
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 182)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fitzames, C., Fisher, C.,
 Bonneau, L., Billault, A., Queller, F., Saurin, W., Bernot, A. and
 Weissbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 182)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000)
COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.
FEATURES
 source
 1..182
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="151719"
 /clone_lib="G"
 /note="Genoscope sequence ID : COAG151CE10DPI-end : T7"
 Location/Qualifiers
 1..182
 58 a 26 c 55 g 39 t 4 others

BASE COUNT
ORIGIN

Query Match 70.0%; Score 18.2; DB 17; Length 182;
 Best Local Similarity 80.0%; Pred. No. 4.7e+02;
 Matches 20; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 1 CGCTTTTCGACATCGTACCTCA 25
 61 CGCTTTTCGACATCGTACCTCA 37

RESULT 7 BB593576 234 bp mRNA linear EST 30-NOV-2000
 BB593576 RIKEN full-length enriched, 4 days neonate male adipose
 Mus musculus cDNA clone B430103F24 5', mRNA sequence.
ACCESSION BB593576.1 GI:11490178
VERSION EST.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 234)
AUTHORS Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P.,
 Hanagaki, T., Hayatsu, N., Hiraoka, T., Hirozane, T., Hodoyama, Y.,
 Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Kono
 Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C.,
 Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shinagawa, A.,
 Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka
 Yoshiaki, A., Yamamura, T., Yasunishi, A., Yoshida, K.,
 Yohsiki, A., Muramatsu, M. and Hayashizaki, Y.
JOURNAL RIKEN Mouse ESTs (Aizawa, K. et al. 2000)
COMMENT Unpublished (2000)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 url: http://genome.gsc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagao, S., Sasaki
 N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Thermostabilization and thermocyclization of thermostable enzymes by
 trehalose and its application for the synthesis of full length

CDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
 Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.

FEATURES
 source
 1..234
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="B430103F24"
 /clone_lib="RIKEN full-length enriched, 4 days neonate
 male adipose"
 /sex="male"
 /tissue="adipose"
 /dev_stage="4 days neonate"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5',
 GAGAGAGAGAGAGATCGATCGATCTTTTCTTTTCTTTTCTTTT 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 10.0 and subtraction to Rot = 229.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5' GAGAGAGAGATCGATCGATCTTTTCTTTTCTTTTCTTTT 3'].
 cDNA was cleaved with XhoI and BamHI. Vector: a
 modified pBluescript KS(+) after bulk excision from Lambda
 FLC 1." 48 a 50 c 84 g 52 t

BASE COUNT
ORIGIN

Query Match 70.0%; Score 18.2; DB 10; Length 234;
 Best Local Similarity 87.0%; Pred. No. 5e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 4 CTTTCGACATCGTACCTCAAT 26
 52 CTTTCGACATCGTACCTCAAT 30

RESULT 8 BB600933 242 bp mRNA linear EST 01-DEC-2000
 BB600933 RIKEN full-length enriched, 13 days embryo lung Mus
 musculus cDNA clone D43002D01 5', mRNA sequence.
ACCESSION BB600933
VERSION BB600933.1 GI:11509534
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 242)
AUTHORS Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P.,
 Hanagaki, T., Hayatsu, N., Hiraoka, T., Hirozane, T., Hodoyama, Y.,
 Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Kono
 Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C.,
 Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shinagawa, A.,
 Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka
 Yoshiaki, A., Yamamura, T., Yasunishi, A., Yoshida, K.,
 Yohsiki, A., Muramatsu, M. and Hayashizaki, Y.

FEATURES	SOURCE
BASE COUNT	63 a 178 c 107 g 87 t
ORIGIN	
Query Match	69.2%; Score 18; DB 12; Length 435;
Best Local Similarity	80.8%; Pred. No. 6.8e+02;
Matches	21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
1 CGTCTTTCGCAGATCGGTACTCAAT	26
219 CGTCTTTCGCAGATCGGTACTCAAT	244
BASE COUNT	98 a 108 c 100 g 103 t 2 others
ORIGIN	
Query Match	69.2%; Score 18; DB 17; Length 411;
Best Local Similarity	80.8%; Pred. No. 6.8e+02;
Matches	21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
1 CGTCTTTCGCAGATCGGTACTCAAT	26
57 CGTCTTTCGCAGATCGGTACTCAAT	32
RESULT 13	
LOCUS	BF588219
DEFINITION	BF588219 435 bp mRNA linear EST 12-DEC-2000
ACCESSION	F01.33.B06.B1.A003 Floral-Induced Meristem 1 (F01) Sorghum
VERSION	BF588219
KEYWORDS	EST.
SOURCE	Sorghum prolinquum.
ORGANISM	Sorghum prolinquum.
REFERENCE	Enkaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
AUTHORS	1 (bases 1 to 435) Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt,L.H.
TITLE	An EST database from Sorghum: floral-induced meristems
JOURNAL	Unpublished (2000)
COMMENT	Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below phred quality 16. The threshold for highest quality sequence is 20. Seq primer: JEN REV High quality sequence stop: 434 POLYA-No.
FEATURES	
SOURCE	Location/Qualifiers 1..435 /organism="Sorghum prolinquum" /db_xref="taxon:132711" /clone_id="Floral-Induced Meristem 1 (F01)" /note="Organ: sperm; Vector: pBel0BAC11; BAC Clones in E-Coli DH10B" phnscript ii from lambda zap ii; Site.1: XhoI; Site.2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested The library was made from poly-A RNA in the cloning vector lambda zap ii. Clones to be sequenced were prepared by mass excision."

RESULT 14
 A0255220 701 bp DNA linear GSS 23-OCT-1998
 LOCUS mgx0009D16r CUGI Rice Blast BAC Library Magnaporthe grisea genomic
 DEFINITION clone mgx0009D16r, DNA sequence.
 ACCESSION A0255220
 VERSION A0255220.1 GI:3779535
 KEYWORDS GSS.
 SOURCE Magnaporthe grisea.
 ORGANISM Magnaporthe grisea
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetes Incertae sedis; Magnaporthaceae; Magnaporthe.
 REFERENCE 1 (bases 1 to 701)
 AUTHORS Yu, Y., Zhu, H., Boyd, C.A., Gaudette, B., Gayle, A., Kingsbury, R.,
 Phillips, R., Sasnowski, M., Wing, R.A. and Dean, R.A.
 TITLE A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
 Genome
 JOURNAL Unpublished (1998)
 COMMENT Contact: Dean RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson University, Clemson, SC 29634
 Tel: 864 656 5737
 Fax: 864 656 4293
 Email: rdeane@clemson.edu
 Seq primer: GGAAACACCTATGACCATG
 Class: BAC ends
 High quality sequence stop: 445.
 Location/Qualifiers
 1..701
 /organism="Magnaporthe grisea"
 /strain="70-15"
 /db_xref="taxon:148305"
 /clone="mgx0009D16r"
 /clone_lib="CUGI Rice Blast BAC Library"
 /tissue_type="Protoplasts"
 /lab_host="E. coli DH10B"
 /note="Vector: pBACWICH; Site_1: HindIII; Site_2: HindIII;
 Rice blast is one of the most devastating fungal diseases
 of rice world wide. It is a filamentous ascomycete with
 a haploid genome (n=7) of approximately 40 Mbp. Rice
 blast is an important model fungal pathogen for studying
 numerous aspects of the fungal-host interaction. In
 order to facilitate genome wide analysis, a BAC library
 containing 19216 clones with an average insert size of 130
 kbp was constructed. This library represents greater
 than 25x genome coverage. High density colony filters
 are available upon request."
 BASE COUNT 168 a 156 c 170 g 206 t 1 others
 ORIGIN
 Query Match 69.2%; Score 18; DB 17; Length 701;
 Best Local Similarity 100.0%; Pred. No. 7.6e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 9 GCAGATCGTACCTCAAT 26
 Db 405 GCAGATCGTACCTCAAT 422
 RESULT 15
 BF757321 150 bp mRNA linear EST 12-JAN-2001
 LOCUS MRO-CT0452-041100-303-h10 CT0452 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF757321
 ACCESSION BF757321
 VERSION BF757321.1 GI:12105221
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 150)

AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W.-Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,
 Brunstein, A.S., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MROct2-MRO-CT0452-
 041100-303-h10&t3-2000-11-04&t4-1)
 Seq primer: puc 18 forward
 High quality sequence start: 13
 High quality sequence stop: 150.
 Location/Qualifiers
 1..150
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CT0452"
 /dev_stage="Adult"
 /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
 SmaI; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No. 196
 ,776 - Ludwig Institute for Cancer Research) profiles
 into the puc 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."
 BASE COUNT 32 a 40 c 46 g 32 t
 ORIGIN
 Query Match 67.7%; Score 17.6; DB 12; Length 150;
 Best Local Similarity 83.3%; Pred. No. 8.5e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 2 GTCTTCGACATCGGACCTCAA 25
 Db 89 GTCTTCGACATCGGACCTCAA 66

Search completed: July 11, 2003, 02:26:05
 Job time : 474.474 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 242.625 Seconds

(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631B-9

Perfect score: 28

Sequence: 1 cgtccgtaggagtagcgcgcgtatgtgtt 28

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues 4109280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rtd:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.2	72.1	193862	2 AC125937	AC125937 Rattus no
2	20	71.4	10100	1 AE005060	AE005060 Haldobacte
3	19.6	70.0	187487	2 AC119388	AC119388 Rattus no
4	19.6	70.0	191844	2 AC122410	AC122410 Mus muscu
5	19.2	68.6	34388	3 CER2888	292813 Caenorhabdi
6	19	67.9	1841	3 DDIGP100	L04286 Dictyostell
7	19	67.9	11884	1 AE004745	AE004745 Pseudomon
8	19	67.9	14687	1 AC126800	AC126800 Mus muscu
9	19	67.9	80015	9 AC110085	AC110085 Homo sapi
10	19	67.9	115045	9 AC004414	AC004414 Homo sapi
11	19	67.9	123770	2 AC120560	AC120560 Rattus no
12	19	67.9	153987	2 AC094593	AC094593 Rattus no
13	19	67.9	179572	2 AC117647	AC117647 Mus muscu
14	19	67.9	189086	2 AC009921	AC009921 Homo sapi
15	19	67.9	209743	2 AL354896	AL354896 Human DNA
16	19	67.9	276248	2 AC087561	AC087561 Mus muscu
17	18.8	67.1	120033	2 AC122725	AC122725 Medicago
18	18.6	66.4	90052	2 AL139429	AL139429 Human DNA
19	18.6	66.4	110120	2 AC117287	AC117287 Rattus no
20	18.6	66.4	129296	9 AC104123	AC104123 Homo sapi
21	18.6	66.4	129667	2 AC095300	AC095300 Rattus no
22	18.6	66.4	143783	2 AC119005	AC119005 Rattus no
23	18.6	66.4	145514	9 AC099509	AC099509 Homo sapi
24	18.6	66.4	154699	2 AC115657	AC115657 Rattus no
25	18.6	66.4	174342	2 AC120962	AC120962 Rattus no
26	18.6	66.4	184061	2 AC121189	AC121189 Rattus no
27	18.6	66.4	190212	2 AC127097	AC127097 Rattus no
28	18.6	66.4	191597	2 AC121956	AC121956 Mus muscu
29	18.6	66.4	191992	2 AC097815	AC097815 Rattus no
30	18.6	66.4	233488	2 AC122228	AC122228 Mus muscu
31	18.6	66.4	240050	1 AL627267	AL627267 Salmonell
32	18.4	65.7	795	3 AF291442	AF291442 Agestiaspi
33	18.4	65.7	801	3 AF291441	AF291441 Agestiaspi
34	18.4	65.7	2578	3 AY051922	AY051922 Drosophil
35	18.4	65.7	4218	1 AF262949	AF262949 Mycobacte
36	18.4	65.7	66497	2 AC015388	AC015388 Drosophil
37	18.4	65.7	86581	2 AC016426	AC016426 Homo sapi
38	18.4	65.7	97146	8 AC007259	AC007259 Arabidops
39	18.4	65.7	110773	9 AC012601	AC012601 Homo sapi
40	18.4	65.7	141867	9 AL450445	AL450445 Human DNA
41	18.4	65.7	149980	2 AP003612	AP003612 Oryza sat
42	18.4	65.7	151386	2 AC013312	AC013312 Homo sapi
43	18.4	65.7	151386	2 AC013312	AC013312 Homo sapi
44	18.4	65.7	158174	2 AC123065	AC123065 Mus muscu
45	18.4	65.7	158923	9 AC097377	AC097377 Homo sapi

ALIGNMENTS

RESULT 1

AC125937

LOCUS: AC125937

DEFINITION: Rattus norvegicus clone CH230-11B2, *** SEQUENCING IN PROGRESS ***

ACCESSION: AC125937.1 GI:21671543

VERSION: AC125937.1 GI:21671543

KEYWORDS: HTG: HTGS_PHASE1.

SOURCE: Norway rat.

ORGANISM: Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

1 (bases 1 to 193862)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsdorfs, S.L., Amarantunge, H.C., Are, J.R., Ayala, M., Banks, T.,
 Barabara, J., Benton, J., Blum, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowler, S., Brileva, M., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Butrell, K.L., Byrd, N.C.,
 Caron, T.E., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Galis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hognes, M., Hollway, C., Hollins, B.,
 Homsl, F., Howard, S., Huber, J., Huylk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karason, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratoch, J., Kueshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawlin, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunolu, G.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunolu, G.,
 Orangunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoshart, N., Sisson, H.,
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weissstock, G., and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 193862)
 Worley, K.C.

Direct Submission
 Submitted (02-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 193862)
 Worley, K.C.

Direct Submission
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: GDM
 Center clone name: CH230-11B2
 Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 143063 bases at least Q40
 Consensus quality: 151194 bases at least Q30
 Consensus quality: 158097 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 57 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 1225
 1325
 2582
 2681
 4358
 4359
 4459
 6004
 6104
 7746
 7846
 9226
 9326
 10350
 10450
 12259
 12359
 13642
 13741
 14870
 14970
 16977
 17077
 18784
 18884
 20889
 20999
 22890
 22990
 24827
 24927
 24938
 25993
 26093
 27825
 27925
 29821
 29921
 32022
 32023
 32123
 34089
 34189
 36125
 36225
 37934
 38034
 39555
 39656
 41719
 41819
 43482
 43582
 46431
 46531
 49047
 49147
 51174
 51274
 55844
 55944
 55945
 58237
 58337
 60907
 61007
 61008
 64724
 64824
 67934
 68034

1224: contig of 1224 bp in length
 1324: gap of unknown length
 2581: contig of 1257 bp in length
 2681: gap of unknown length
 4358: contig of 1677 bp in length
 4359: gap of unknown length
 4459: contig of 1545 bp in length
 6004: gap of unknown length
 6104: contig of 1642 bp in length
 7746: gap of unknown length
 7846: contig of 1380 bp in length
 9226: gap of unknown length
 9326: contig of 1024 bp in length
 10349: gap of unknown length
 10449: gap of unknown length
 12258: contig of 1809 bp in length
 12358: gap of unknown length
 13641: contig of 1283 bp in length
 13741: gap of unknown length
 14869: contig of 1128 bp in length
 14969: gap of unknown length
 16976: contig of 2007 bp in length
 17076: gap of unknown length
 18783: contig of 1707 bp in length
 18883: gap of unknown length
 20898: contig of 2015 bp in length
 20998: gap of unknown length
 22889: contig of 1891 bp in length
 22989: gap of unknown length
 24827: contig of 1838 bp in length
 24927: gap of unknown length
 25993: contig of 1066 bp in length
 26093: gap of unknown length
 27824: contig of 1731 bp in length
 27924: gap of unknown length
 29821: contig of 1897 bp in length
 29921: gap of unknown length
 32022: contig of 2101 bp in length
 32122: gap of unknown length
 34088: contig of 1966 bp in length
 34188: gap of unknown length
 36124: gap of unknown length
 36224: gap of unknown length
 37933: contig of 1709 bp in length
 38033: gap of unknown length
 39555: contig of 1522 bp in length
 39656: gap of unknown length
 41718: contig of 2063 bp in length
 41818: gap of unknown length
 43481: contig of 1663 bp in length
 43581: gap of unknown length
 46430: contig of 2849 bp in length
 46530: gap of unknown length
 49047: contig of 2517 bp in length
 49147: gap of unknown length
 51174: contig of 2027 bp in length
 51274: gap of unknown length
 55844: contig of 4570 bp in length
 55944: gap of unknown length
 55945: contig of 2293 bp in length
 58237: gap of unknown length
 58337: gap of unknown length
 60907: contig of 2570 bp in length
 61007: gap of unknown length
 61008: contig of 3716 bp in length
 64724: gap of unknown length
 64824: contig of 3110 bp in length
 67934: gap of unknown length
 68034: contig of 2310 bp in length


```

gene
365. .2643
/gene="VNG1426h"
365. .2643
/gene="VNG1426h"
/note="hypothetical protein"
/codon_start=1
/product="Vng1426h"
/protein_id="AAG19740.1"
/db_xref="GI:10580928"
/translation="MMDLTFQFDLLVAVAGLDEPRGAIKDELQYYSSEIHGRLY
PMDLTVKGLIDKDELQRTNYTLTTRCTRELEAREMETQYVAGS"
2647. .2889
/gene="VNG1427h"
2647. .2889
/gene="VNG1427h"
/note="hypothetical protein"
/codon_start=1
/product="Vng1427h"
/protein_id="AAG19741.1"
/db_xref="GI:10580929"
/translation="MRAKATATANGVAGSLAVSVVAVQFDLPVFLVAVPPELFR
RREGSSPRTKRCPECGFTTKRPSVHCHDGTELD"
2647. .4034
/gene="hclA"
2647. .4034
/complement(2886. .4034)
/note="hclA"
/codon_start=1
/product="Hcl-like protein"
/protein_id="AAG19742.1"
/db_xref="GI:10580930"
/translation="MGDTRIDGRRAAMAIQTITLALVYVYVTEVGTGFLFYLA
TRPVRRVDRHRSSSSVAVVSLTALPVVLSYTTAVGIDELTRVLPQIDLT
AVLAVOPVYDSRANQDQSLGQPSVAAINTLRSAGVAGFLGTAFLFAVIT
FNVLQIVPSSISIPYPTMGLTGAASLIVVGLKIWMIPLLGMALVAAARSGGAGT
FVAVGVAAVAVIPIDLVAVPVYSGDLHGLVLAIVYVPLIMVYIGTFLCPMVL
VEVVFARLVPELVAGEGIAVPAGADDTFASADPESADPDDTA"
4126. .5136
/gene="VNG1429c"
4126. .5136
/gene="VNG1429c"
/note="conserved hypothetical protein"
/codon_start=1
/translation="Vng1429c"
/protein_id="AAG19743.1"
/db_xref="GI:10580931"
/translation="MTFLAGGTGTPLLAGARRVFPATTTVVGNTGDDVALGLLY
CPDLDTVLEGGVLDRETWGLADDSSTHDTYTLAAADIDPPTRYLPDADOTA
GRDIARRRSSASEMFIDGRBAVHTLGRGLIDEGHTLETYRLRLADFDLSDYL
PMSNDPVAITVQTPDGEHQTEFVVAHGDPTVEDVEFGEGRATAOPATEIRKCP
VVVGPSPNTSIGPMALDGLADALRDAOVAVASPRVEDVYSGRAKMAAAGHDS
TAGVAAADVPADAFVLDTASTDLDRVYVTDISLDTEDAEVAVARCRDALVAASGE
VT"
5198. .5869
/gene="VNG1431c"
5198. .5869
/gene="VNG1431c"
/note="conserved hypothetical protein"
/codon_start=1
/translation="Vng1431c"
/protein_id="AAG19744.1"
/db_xref="GI:10580932"
/translation="MSAGEHAGGVFLGTAIDEPREARARAWDRREFLPADVAVM
VDDLSLDAFLRPVAVNRAVATGLRVAAYACAHADILTEVNAHCRQDELCDAGAG
ETLRDTSRLCEVRASDEGAASVAVKAVAGVGHLPVLAQATIERAGAAAIHVAMD

```

```

gene
365. .6992
/gene="dhs"
365. .6992
/note="VNG1432c"
6000. .6992
/gene="dhs"
/note="dhs"
/codon_start=1
/product="deoxyhypusine synthase"
/protein_id="AAG19745.1"
/db_xref="GI:10580933"
/translation="MTGDDADETHNVPSSDEDLPDVRGNYDPSGCEFPFELDSY
ATGFOASHADAVDIPREMERDATTITLTSITVSSGLREVVAHLVRENYDVIT
TSGSITEDIITKARPKRGENDVDEALREGLINRLNINIVPSDRVWLEBYLYDFE
EPFADQKRTPTAFARRELIDDEDSILKNAADNDIPVCPALTDALINFLVYVRO
GYDSEVSGEILDDYDALIEGLADDTGLICVGVGPKHAHIMTNLEFGADVAVYIS
TGMGDSLSGAPPEEAVSMCKIRKEDAEPRYALIEATLVPLVAGAFENP"
7078. .7257
/gene="rps17e"
7078. .7257
/note="VNG1433c"
7078. .7257
/gene="rps17e"
/note="rps17e"
/codon_start=1
/product="30S ribosomal protein S17e"
/protein_id="AAG19746.1"
/db_xref="GI:10580934"
/translation="MAIKPDYVKKTGNIIMERYODAFSREDFEHNKDAVETLNIESK
NVRKRIAGYITHRN"
7304. .8335
/gene="asd"

Query Match 71.4%; Score 20; DB 1; Length 10100;
Best Local Similarity 82.1%; Pred. No. 51;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGTTCGTGGATATGCCGTACGTGTT 28
Db 3946 CGTTCGTGGAGACGTCGTCGTGGGTT 3919

RESULT 3
AC119388
LOCUS
DEFINITION
Rattus norvegicus clone CH230-215C17, *** SEQUENCING IN PROGRESS
ACCESSION
AC119388.6 GI:21902577
VERSION
HTG: HTGS_PHASE1.
KEYWORDS
Rattus norvegicus.
SOURCE
Rattus norvegicus.
ORGANISM
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 187487)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amaralunga,H.C., Ate,J.R., Ayala,M., Banks,T.,
Barbarta,U., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowles,S., Bileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dethorne,S.R., David,R.,
Devaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

```

Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homi, F., Howard, S., Huber, J., Huliyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korval, J., Kovar, C., Kravovic, U., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozdo, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., McElroy, M. P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okunolu, G., Oregany, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Plickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rivers, R., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Taney, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, S., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Unpublished
2 (bases 1 to 187487)
Worley, K. C.

Direct Submission
Submitted (26-Apr-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 187487)
Worley, K. C.

Direct Submission
Submitted (23-Jul-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced gi:20467907.

Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GWKF
Center clone name: CH230-215C17

Summary Statistics

Sequencing vector: Plasmid
Assembly program: Phrap, version 0.990329
Consensus quality: 14644 bases at least Q40
Consensus quality: 153280 bases at least Q30
Consensus quality: 157543 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draht_data.html).
NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1223: contig of 1223 bp in length
1224: gap of unknown length
1324: contig of 1142 bp in length
2465: gap of unknown length
2565: gap of unknown length
3894: contig of 1329 bp in length
3895: gap of unknown length
5712: contig of 1718 bp in length
3995

5713: gap of unknown length
5813: contig of 1000 bp in length
6813: gap of unknown length
8166: contig of 1254 bp in length
8167: gap of unknown length
8266: gap of unknown length
8267: contig of 1135 bp in length
9401: gap of unknown length
9402: gap of unknown length
9502: contig of 1458 bp in length
10959: gap of unknown length
10960: gap of unknown length
11059: gap of unknown length
11060: contig of 1265 bp in length
12325: gap of unknown length
12424: gap of unknown length
12425: contig of 1587 bp in length
14012: gap of unknown length
14111: gap of unknown length
14112: contig of 1666 bp in length
15177: gap of unknown length
15778: gap of unknown length
15878: contig of 2002 bp in length
17880: gap of unknown length
17889: gap of unknown length
17980: contig of 1682 bp in length
19661: gap of unknown length
19662: gap of unknown length
19761: gap of unknown length
21649: contig of 1888 bp in length
21749: gap of unknown length
21750: gap of unknown length
23612: contig of 1863 bp in length
23712: gap of unknown length
23713: contig of 1948 bp in length
25660: gap of unknown length
25760: gap of unknown length
27188: contig of 1428 bp in length
27288: gap of unknown length
27289: gap of unknown length
27289: contig of 2327 bp in length
29715: gap of unknown length
29716: gap of unknown length
32390: contig of 2675 bp in length
32391: gap of unknown length
32490: gap of unknown length
33491: contig of 2562 bp in length
35053: gap of unknown length
35152: gap of unknown length
37301: contig of 2149 bp in length
37302: gap of unknown length
37401: gap of unknown length
37402: gap of unknown length
39449: contig of 2048 bp in length
39450: gap of unknown length
39459: gap of unknown length
41479: contig of 1930 bp in length
41480: gap of unknown length
41579: gap of unknown length
41580: contig of 2791 bp in length
44371: gap of unknown length
44470: gap of unknown length
44471: contig of 2113 bp in length
46583: gap of unknown length
46584: gap of unknown length
46684: contig of 4010 bp in length
50694: gap of unknown length
50793: gap of unknown length
50994: gap of unknown length
53466: contig of 2673 bp in length
53467: gap of unknown length
53566: gap of unknown length
53567: contig of 2775 bp in length
56341: gap of unknown length
56441: gap of unknown length
58665: contig of 2224 bp in length
58765: gap of unknown length
61150: contig of 2385 bp in length
61151: gap of unknown length
61230: gap of unknown length
61250: contig of 2350 bp in length
61251: gap of unknown length
63601: gap of unknown length
63700: gap of unknown length
65100: contig of 2400 bp in length
66101: gap of unknown length
66200: gap of unknown length
68427: contig of 2227 bp in length
68428: gap of unknown length
68527: gap of unknown length
71553: contig of 3026 bp in length
71554: gap of unknown length
71653: gap of unknown length
71654: contig of 2662 bp in length
74316: gap of unknown length
74415: gap of unknown length
77732: contig of 3317 bp in length
77733: gap of unknown length
77832: gap of unknown length
80822: contig of 2990 bp in length
80823: gap of unknown length
80922: gap of unknown length
84039: contig of 3117 bp in length
84139: gap of unknown length
84140: gap of unknown length
87817: contig of 3678 bp in length
87818: gap of unknown length
87917: gap of unknown length
91189: contig of 3272 bp in length
91189: gap of unknown length
91289: gap of unknown length

```

* 91290 96104: contig of 4815 bp in length
* 96105 96204: gap of unknown length
* 96205 101241: contig of 5037 bp in length
* 101342 101341: gap of unknown length
* 101342 106604: contig of 5263 bp in length
* 106605 106704: gap of unknown length
* 106705 111595: contig of 4891 bp in length
* 111596 111595: gap of unknown length
* 115757 115757: contig of 4062 bp in length
* 115758 115857: gap of unknown length
* 115858 121880: contig of 6023 bp in length
* 121881 121880: gap of unknown length
* 121981 126071: contig of 4091 bp in length
* 126072 126171: gap of unknown length
* 133330 133329: contig of 7158 bp in length
* 133430 138831: gap of unknown length
* 138832 138931: gap of 5402 bp in length
* 138932 146738: gap of unknown length
* 146739 146838: contig of 7807 bp in length
* 146839 155332: gap of unknown length
* 155333 155432: contig of 8494 bp in length
* 155433 164997: gap of unknown length
* 164998 165097: contig of 9565 bp in length
* 165098 174999: gap of unknown length
* 175000 175099: contig of 9902 bp in length
* 175100 187487: gap of unknown length

```

```

Query Match
Best Local Similarity 84.6%; DB 2; Length 187487;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

3 TTCTGCGAGATGCGTCATGCTGT 28
51506 TTCTGCGAGATGCGTCATGCTGT 51531

```

```

RESULT 4
AC122410 191844 bp DNA linear HTG 05-JUL-2002
LOCUS Mus musculus chromosome UNK clone RP24-142A1, WORKING DRAFT
DEFINITION AC122410
AC122410
AC122410.2 GI:21699728
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE 1 (bases 1 to 191844)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 191844)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Jul 5, 2002 this sequence version replaced gi:21105870.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
Project Information
Center project name: M.BR0142A01

```

```

----- Summary Statistics -----
Sequencing vector: MJ3.0%
Sequencing vector: plasmid: 100%
Chemistry: Dye primer ET; 0% of reads
Chemistry: Dye terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 188883 bases at least Q40
Consensus quality: 189311 bases at least Q40
Consensus quality: 189743 bases at least Q20
Insert size: 193000; agarose-IP
Insert size: 192017; sum-of-contigs
Quality coverage: 0.00 in Q20 bases; agarose-IP
Quality coverage: 11.99 in Q20 bases; sum-of-contigs

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1185: contig of 1185 bp in length
* 1186 1285: gap of unknown length
* 1286 2466: contig of 1181 bp in length
* 2467 2566: gap of unknown length
* 2567 8024: contig of 5458 bp in length
* 8025 8124: gap of unknown length
* 8125 18649: contig of 10525 bp in length
* 18650 18749: gap of unknown length
* 18750 42353: contig of 23604 bp in length
* 42354 42453: gap of unknown length
* 42454 82772: contig of 40319 bp in length
* 82773 82872: gap of unknown length
* 82873 191197: contig of 108325 bp in length
* 191198 191297: gap of unknown length
* 191298 191844: contig of 547 bp in length.

```

```

FEATURES
source
1. 191844
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-142A1"
1. 1185

```

```

misc_feature
/note="assembly_name:Contig22"
1286..2466
/note="assembly_name:Contig23"
2367..8024
/note="assembly_name:Contig24"
8125..18649
/note="assembly_name:Contig25"
18750..42353
/note="assembly_name:Contig26"
42454..82772
/note="assembly_name:Contig27"
82873..191197
/note="assembly_name:Contig28"
clone_end:sp6
vector_side:left"
191298..191844
/note="assembly_name:Contig8"
BASE COUNT 57770 a 38253 c 38232 g 56888 t 701 others
ORIGIN

```

```

Query Match
Best Local Similarity 70.0%; Score 19.6; DB 2; Length 191844;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 GTTCTGCGAGATGCGTCATGCTGT 27
DB 153158 GTTCTGCGAGATGCGTCATGCTGT 153183

```


TITLE .

Smith, K., Spencer, D., Wong, G.K., Wu, Z.
complete genome sequence of *Pseudomonas*

III, K. 1

opportunistic pathogen
JOURNAL Nature 406 (6799), 959-964 (2000)
MEDLINE 20437337
REFERENCE 10984043
AUTHORS 2 (bases 1 to 11884)
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hutnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Madman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Lardig,K., Lim,R.M., Smith,K.A., Spencer,D.H., Wong,G.K.S., Wu,Z., Paulsen,I.T., Kelder,J., Sater,M.H., Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
TITLE Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
FEATURES
Source Location/Qualifiers
1. 11884
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
102. 419
/gene="PA3216"
102. 419
/gene="PA3216"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06604.1"
/db_xref="GI:9949337"
/translation="MSSPDERFHSFAEFPYLYLOESHNAVCRLHYVSLVLAVLE HAVASQOMLMLALPLVGYGPAWGHVEFEKNRPAATFKRYPLMSFGMDVMKDAFTGR IR"
gene 636. 2027
/gene="PA3217"
636. 2027
/gene="PA3217"
/codon_start=1
/transl_table=11
/product="probable adenylate cyclase"
/protein_id="AAG06605.1"
/db_xref="GI:9949338"
/translation="MKPLLDPDRSPTSRSVSMREYSRVLIYIACGASIAAGTYQF SYGLIMVAYALLPLPHLYHGORHODPKRVTRALLAVDAVHCGMALIGSVYP SIMLVLSEFALYIGILRILGALIVASASALLAVIVAPILGNTPEVAASITF CGLYCITAFPGHOOGRLAQRQEIARBOEKARLARLNIAKITSPOWEITEFSKVS VRLTQKKLTFEFSIDIRGFELSEELAEALITLNNLYNEMSKIALKGTIDKTV GDCVAVFFGDPSTOGAKKDAVAASGIMAKRHVKLRQOWRAQGITKPIEIRGINT GCTVGNFGADTRMDYTLIGREVNILASRLSESAGETILISHETYSLIKDIIMCRDG QIAVGFSPVOIYQVVDSDRDLGAAPSYVEHLPJFSMYLDTNNIONYDERVIAL QOAEERLRDQVIL"
gene 2569. 3042
/gene="PA3218"
2569. 3042
/gene="PA3218"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06606.1"
/db_xref="GI:9949339"
/translation="MEQFGIDALAPQVLEAPAEELVAQARGAHOHRPAPMEGAOE GIAGGQRPRTAVOVFGKARMDRGEVOALAPAPRGPAPERPLGGOVHIREGID OPAEAPVPGGDRIRAGTGOAAEVRGRHLYVAQGFALQGOAGDSDVDLR"
CDS complement(3069..3881)
/gene="PA3219"
complement(3069..3881)
/gene="PA3219"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06607.1"
/db_xref="GI:9949340"
/translation="MSTAQRFSTNRKQRVFTLWISDVHLGTRDCAEHLAFLKRYHA

DRIVAGDIIIDGWLKRGIIYPOAHNTVIRLLTLKRGTEVIYVTGNHDEFLRRSS LIIENIRLVDEAVSTADGRKLLVIHGGDFPVITRYHRLWLAFLGDSAYEFTLLNRL NHRRRNGCYWNSLSALKHKVKTAVNISDFEERIALHECHKRLDGVYCHIHAEI RRIEGVIDLNGDWESCTALIEHWDISIQILRAIEQARLAQAAAEVPA"
gene 4028. 4792
/gene="PA3220"
4028. 4792
/gene="PA3220"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="AAG06608.1"
/db_xref="GI:9949341"
/translation="MGSITSLRHYSHDOYVSHDQVLGSGCLDPEVGRGSRVL RQFAVYPAQARHACSSPSGRCLVLDPCGSLIGGEHAKGRLEFEPQALSTP PAQOLVNLASGPIINDPVIAHQAGLLASLASERQPPROPPLPLASLDATYDRHA AHPQVADLARIAGLSVAREFSLTFLAEFGQPMQYAKRSRLHQAEHLLESSLAVGEI AARYVASQSAFTALVAREFGVTPROLRRRESRDTG"
CDS 4857. 5192
/gene="csaA"
/note="PA3221"
4857. 5192
/gene="csaA"
/codon_start=1
/transl_table=11
/product="CsaA protein"
/protein_id="AAG06609.1"
/db_xref="GI:9949342"
/translation="MDIEMODFERVELRVGTILISAAPNPKARPAYVLEVDLPLGL RTSSAQLTAHQAEQLVGRQVLCVNCNAPKQIAGVRSVLTGVYDSEDKVYLAGFQK PLNGARLA"
gene 5202. 6101
/gene="PA3222"
5202. 6101
/gene="PA3222"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06610.1"
/db_xref="GI:9949343"
/translation="MNRSAALALHGLLFGTLGVFGKILASAPGYIVFGRALPAVL ALGLFARLAGQAMRALGPQAWMLRLALGGLLGHVWSEFFIAVKTAGVAILVGFASPE AFVYILBELFERERIHAEGLLVAVLGVLPNEDLAGATLGLMALISGLLEFA LLSIANNSGQGVQVQALQONVVALCLLPFVSELPALPDMLMIALLGCGCG LASHLSFVASLAVVAPRAAIYFALEPIYGIAPMLLPHETGPPMFLGALIIIAIYL SARIGARKTRAVDERQDPERA"
CDS complement(6206..6847)
/gene="acpD"
/note="PA3223"
complement(6206..6847)
/gene="acpD"
/codon_start=1
/transl_table=11
/product="acyl carrier protein phosphodiesterase"
/protein_id="AAG06611.1"
/db_xref="GI:9949344"
/translation="MSRVIVTESSARQGSVRLTAFTISHMKIAPADQFOYRDLA REPLPHDELILGAWTTPCDGHSAAERAAERSRNRIEELMAADVILAPMTFAPL SSLSKSWFDHVLRAGLTFRYAQGEGLQGRARVLTARGGIVAGGGLDQHEPYLRQV LGFVGIDHVFPIHAEGNMMPGEFEKEKGLARAREMRQALFETDISLCPVLPFLR"
gene complement(6994..7311)
/gene="PA3224"
complement(6994..7311)
/gene="PA3224"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06612.1"
/db_xref="GI:9949345"
/translation="MKRIICCPALSLPLSAPFAYIEVDPOLNGTEVSYTODIERDKG AILLTNNGQAAECTAVFRNGPELPKRVKGVIOGQNSNFTAPGGRDIIKRIKLTCK IAK"
gene 7511. 8440

CDS

/gene="PA3225"
7511..8440
/gene="PA3225"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="AA006614.1"
/db_xref="GI:9949346"
/translation="MKAPRTLDQMTQAVVHDGGRQAQAAELHNSQSSISYTVARM
QEOGLVPLRIDGRKAVITRAGCYLLRSKQLVKAQGLELAHMEQGEPEVRLV
DAAYPTARIYSLSAPFQSGCVLLREYLSGVEEALDQGAADALISGLIPGILG
ADLSVFEVAVAHDPPLHRLQRELTQHDDETQVYVIRDSGLQPRDHGMGADERM
TVGSLATAFVVGGLFAMLPRIITEREQSQLPPLPSQGVQNSRYLYTNKEK
PLGPSQIIMEMKSFANVPLNAPFAPEPAE"
8447..9274
/gene="PA3226"
8447..9274
/gene="PA3226"
/codon_start=1
/transl_table=11
/product="probable hydrolase"
/protein_id="AA006614.1"
/db_xref="GI:9949347"
/translation="MPDFQRODCTIYDDHDPGTERAPVLLHGLSSARDWEYOLPA
LIGRYRLVPLDNGHRSKPGCYSMAGFAEDCALDLRLCGVPLVVGISMGCMIG
FOLACRPDLIRSLITVNSTPEYTPKPREHELVAKRKLISLRTIGRALGILF
PKAEQGLIKRKLEARNENDKRAVLSADALIGVQVERLAIACPTLVISADRYTP
VSLKQATCERILDLVVEDSRHATPMDQPIFTTLLGLADIKKEQCSVVS"
9256..9819
/gene="p1A"
/note="PA3227"
9256..9819
/gene="p1A"
/codon_start=1
/transl_table=11
/product="peptidyl-prolyl cis-trans isomerase A"
/protein_id="AA006615.1"

gene

CDS

Query Match

Best local similarity 67.9% Score 19; DB 1; Length 11884;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GTTCGTGGAGATGCTGCATGCTGT 28
1457 GTTCGTGGAGATGCTGCATGCTGT 1483

RESULT 8

AC126800 14687 bp DNA linear HTG 09-JUL-2002
LOCUS AC126800/c
DEFINITION Mus musculus chromosome UNK clone RP24-447D19, WORKING DRAFT
ACCESSION AC126800
VERSION AC126800.1 GI:21717270
KEYWORDS HTG; HTGS; PHASE1; HTGS; DRAFT.
SOURCE house mouse
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 14687)
McPherson, J.D. and Waterston, R.H.
The sequence of Mus musculus clone
21 (bases 1 to 14687)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (09-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

COMMENT

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml

Contact: submissions@wustl.edu
Project information
Center project name: W_BB0447D19

Summary Statistics

Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 11885 bases at least Q40
Consensus quality: 12505 bases at least Q30
Consensus quality: 13093 bases at least Q20
Insert size: 175000; agarose-fp
Insert size: 13787; sum-of-coverage
Quality coverage: 0.12 in Q20 bases; sum-of-coverage
Quality coverage: 1.48 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently
consists of 10 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1121: contig of 1121 bp in length
1122 1221: gap of unknown length
1222 2542: contig of 1321 bp in length
2543 2642: gap of unknown length
2643 3696: contig of 1054 bp in length
3697 3797: gap of unknown length
3797 5103: contig of 1307 bp in length
5104 5203: gap of unknown length
5204 6395: contig of 1192 bp in length
6396 6495: gap of unknown length
6496 8013: contig of 1518 bp in length
8014 8113: gap of unknown length
8114 10012: contig of 1899 bp in length
10013 10112: gap of unknown length
10113 11603: contig of 1491 bp in length
11604 11703: gap of unknown length
11704 13134: contig of 1431 bp in length
13135 13235: gap of unknown length
13235 14687: contig of 1453 bp in length.

FEATURES

source

1.14687
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="7UNK"
/clone="RP24-447D19"
1.1121
/note="assembly_name:Contig15"
1222.2542
/note="assembly_name:Contig31"
2643.3696
/note="assembly_name:Contig33"
3797.5103
/note="assembly_name:Contig35"
5204.6395
/note="assembly_name:Contig38"
6496.8013
/note="assembly_name:Contig40"
8114.10012
/note="assembly_name:Contig41"
10113.11603
/note="assembly_name:Contig42"
11704.13134
/note="assembly_name:Contig43"
13235.14687
/note="assembly_name:Contig44"
misc_count 3528 a 3262 c 3191 g 3797 t 909 others
BASE COUNT
ORIGIN


```

repeat_region      /rpt_family="AT_rich"
                    14527..14866
repeat_region      /rpt_family="ERV1"
                    14867..14965
repeat_region      /rpt_family="ERV1"
                    16820..17126
repeat_region      /rpt_family="Alu"
                    18043..18394
repeat_region      /rpt_family="ERV1"
                    18412..18744
repeat_region      /rpt_family="ERV1"
                    18745..18789
repeat_region      /rpt_family="ERV1"
                    18836..18900
repeat_region      /rpt_family="TTCGCCn"
                    18901..18966
repeat_region      /rpt_family="ERV1"
                    18968..19124
repeat_region      /rpt_family="MERL_type"
                    20267..20317
repeat_region      /rpt_family="AT_rich"
                    20430..20538
repeat_region      /rpt_family="AT_rich"
                    21037..21057
repeat_region      /rpt_family="AT_rich"
                    21576..22017
repeat_region      /rpt_family="L2"
                    22184..22548
repeat_region      /rpt_family="MALR"
                    22561..22868
repeat_region      /rpt_family="MALR"
                    24085..24256
repeat_region      /rpt_family="L2"
                    24283..24402
repeat_region      /rpt_family="L1"
                    24404..24491
repeat_region      /rpt_family="MERS3"
                    24503..25061
repeat_region      /rpt_family="L2"
                    25067..25137
repeat_region      /rpt_family="L2"
                    25166..26022
repeat_region      /rpt_family="L2"
                    26239..26260
repeat_region      /rpt_family="AT_rich"
                    26472..26500
repeat_region      /rpt_family="AT_rich"
                    26870..26898
repeat_region      /rpt_family="AT_rich"
                    27957..28030
repeat_region      /rpt_family="AT_rich"
                    28166..28189
repeat_region      /rpt_family="AT_rich"
                    28670..28714
repeat_region      /rpt_family="AT_rich"
                    28745..29121
repeat_region      /rpt_family="L2"
                    29265..29563
repeat_region      /rpt_family="ERV1"
                    29643..29778
repeat_region      /rpt_family="MIR"
                    29813..29856
repeat_region      /rpt_family="AT_rich"
                    30211..30507
repeat_region      /rpt_family="Alu"
                    30839..31786

```

Query Match 67.9%; Score 19; DB 9; Length 80015;
 Best Local Similarity 81.5%; Pred. No. 1.9e+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

2 GTTCGTGGATAGTCGCTCATGTT 28
 ||||| ||||| |||||

Db 54225 GTTCGTGGATAGTCGCTCATGTT 54199

RESULT 10
 AC004414/c 115045 bp DNA linear PRI 21-DEC-1999
 LOCUS Homo sapiens PAC clone RP4-724E13 from 7p11.2-p12, complete
 DEFINITION
 ACCESSION AC004414
 VERSION AC004414.1 GI:2979592
 KEYWORDS HMG
 SOURCE Homo sapiens
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 TITLE (bases 1 to 115045)

JOURNAL
 TITLE The sequence of Homo sapiens PAC clone RP4-724E13
 AUTHORS Smith A and Le T
 REFERENCE
 TITLE Unpublished
 AUTHORS 2 (bases 1 to 115045)
 JOURNAL Waterston, R.

REFERENCE
 TITLE Direct Submission
 AUTHORS Submitted (20-MAR-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 JOURNAL 3 (bases 1 to 115045)
 REFERENCE
 TITLE Waterston, R.
 JOURNAL Direct Submission

COMMENT
 TITLE Submitted (21-DEC-1999) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 JOURNAL
 TITLE Genome Center
 REFERENCE
 TITLE Center: Washington University Genome Sequencing Center
 AUTHORS
 TITLE Web site: http://genome.wustl.edu/gsc
 REFERENCE
 TITLE Contact: sapiens@watson.wustl.edu
 JOURNAL
 TITLE Summary Statistics
 REFERENCE
 TITLE Center project name: H_DJ0724E13

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry; an attempt was made to resolve all sequencing problems,
 such as compressions and repeats; all regions were covered by
 sequence from more than one subclone; and the assembly was
 confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and
 sequencing collaboration between the NHGRI Chromosome 7 Mapping
 Project (Eric D. Green, Director), John D. McPherson in the
 Department of Genetics (Washington University), and the Washington
 University Genome Sequencing Center. For additional information
 about the map position of this sequence, see
 http://www.nhgri.nih.gov/DIR/GMB/CH7
 mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
 This clone was derived from human PAC library RPI-4, prepared by
 Pieter de Jong and coworkers at Roswell Park Cancer Institute,
 using the method described by Ioannou et al., Nature Genetics
 6:84-9 (1994). The library is from one male donor. For further
 details, see http://bacpac.med.buffalo.edu/
 The clone is available from Genome Systems, Inc.
 (http://www.genomesystems.com).

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP4-537J23. Actual start of
 this clone is at base position of 1 of RP4-724E13; actual end is at
 115045 of RP4-724E13.

FEATURES This clone contains STS SWSS3034 (NTD:G1233489).
 Location/Qualifiers
 source 1. 115045

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p11.2-p12"
/clone="RP4-724E13"
/clone_id="RPC1-4"
115. 211
repeat_region /rpt_family="MIR"
785. 1150
repeat_region /rpt_family="MALR"
1151. 1286
repeat_region /rpt_family="MER1_type"
1518. 1828
repeat_region /rpt_family="L2"
1845. 2157
repeat_region /rpt_family="Alu"
3529. 3585
repeat_region /rpt_family="MIR"
3862. 3950
repeat_region /rpt_family="MIR"
4004. 4351
repeat_region /rpt_family="MER51A"
4854. 5198
repeat_region /rpt_family="L1"
5197. 5608
repeat_region /rpt_family="L1"
6177. 6227
repeat_region /rpt_family="MER4-group"
6229. 6384
repeat_region /rpt_family="MER4-group"
6397. 6635
repeat_region /rpt_family="MER4-group"
6646. 6738
repeat_region /rpt_family="MER4-group"
7798. 7919
repeat_region /rpt_family="MER4-group"
7937. 8097
repeat_region /rpt_family="MER4-group"
8089. 8568
repeat_region /rpt_family="MER4-group"
8564. 8935
repeat_region /rpt_family="MER4-group"
8937. 9028
repeat_region /rpt_family="MER4-group"
9025. 9461
repeat_region /rpt_family="MER4-group"
9471. 9512
repeat_region /rpt_family="L2"
9600. 9677
repeat_region /rpt_family="L2"
10659. 10961
repeat_region /rpt_family="Alu"
11884. 12003
repeat_region /rpt_family="L1"
12172. 12299
repeat_region /rpt_family="Alu"
12765. 12935
repeat_region /rpt_family="L1"
14936. 15312
repeat_region /rpt_family="MALR"
15680. 16117
repeat_region /rpt_family="MALR"
16252. 16430
repeat_region /rpt_family="L1"
16432. 16727
repeat_region /rpt_family="Alu"
16744. 16809
repeat_region /rpt_family="L1"
16812. 17261
repeat_region

```

```

/rpt_family="MER4-group"
17262. 17711
repeat_region /rpt_family="L1"
17742. 18017
repeat_region /rpt_family="Alu"
18646. 18855
repeat_region /rpt_family="MER1_type"
19218. 19709
repeat_region /rpt_family="L1"
22186. 22273
repeat_region /rpt_family="MIR"
22826. 23017
repeat_region /rpt_family="Retroviral"
23020. 23168
repeat_region /rpt_family="Alu"
23412. 23456
repeat_region /rpt_family="Retroviral"
23491. 23642
repeat_region /rpt_family="MIR"
23820. 24014
repeat_region /rpt_family="MER1_type"
24355. 24609
repeat_region /rpt_family="MIR"
24903. 25536
repeat_region /rpt_family="MER4-group"
26083. 26644
repeat_region /rpt_family="L1"
26646. 26947
repeat_region /rpt_family="Alu"
26979. 27362
repeat_region /rpt_family="L1"
27449. 28151
repeat_region /rpt_family="L1"
28223. 28485
repeat_region /rpt_family="MER2_type"
28487. 29298
repeat_region /rpt_family="MER21_g"
29312. 29978
repeat_region /rpt_family="L1"
29990. 30366
repeat_region /rpt_family="L1"
30351. 32031
repeat_region /rpt_family="L1"
32030. 32521
repeat_region /rpt_family="L1"
32536. 32929
repeat_region /rpt_family="Retroviral"
32937. 33598
repeat_region /rpt_family="L1"
33639. 33939
repeat_region /rpt_family="Alu"
34345. 34628
repeat_region /rpt_family="L1"
35643. 35945
repeat_region /rpt_family="Alu"
35986. 36474
repeat_region /rpt_family="L1"
36497. 36719
repeat_region /rpt_family="L1"

```

Query Match 67.9% Score 19; DB 9; Length 115045;
 Best Local Similarity 81.5% Pred. No. 2e+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 GTTCGTGGATGATCGTATGCTT 28
 DB 41511 GTTCGTGGATGATCGTATGCTT 41485

RESULT 11
 AC120560 123770 bp DNA linear HTG 18-JUL-2002
 LOCUS AC120560/c
 DEFINITION Rattus norvegicus clone CH230-330L5, *** SEQUENCING IN PROGRESS

***, 29 unordered pieces
ACT120560

AC120560.2 GI:21745753

HTG; HTGS_PHASE1.
Norway 2024

Norway rat.
Rattus norvegicus

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE AUTHORS

1 (bases 1 to 123770)
Muzny D W 23-

Alsbrooks, S.L., Amaratunge, H.C., Alvarado, R., Alt, Osman, E.R., Allen, C.,
Barbarta, J., Benton, J., Blinmeyer, K., Blinmeyer, K., Blinmeyer, K., Banks, T.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.D.,
Buhay, C., Butch, P., Burkett, C., Burrell, K.L., Byrd, N.J.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
Cleaveland, C.D., Cox, C., Coyle, M.D., Dahorne, S.R., David, R.,
DeValle, M.L., Davis, C., Delgado, O., Denn, A.L., Ding, I., Dinh, H.H.,
DeLanay, K.R., L., Draper, H., Dugan-Rocha, S., Durbin, J.,
Eamheart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flaigg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorell, D.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hayes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hoque, M., Holloway, C., Hollins, B.,
Homsli, F., Jacobson, B., Jia, Y., Huber, J., Hui, S., Hume, J., Jackson, L.E.,
Karlsson, E., Kelly, S., Khan, U., King, L., Koryah, J., Kovar, C.,
Kratavich, C., Kureishi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, H.,
Li, J., Li, Z., Lichstein, O., Lien, C., Liu, J., Liu, W., Louisguez, R.,
Lopez, R.J., Lu, X., Luchter, A., Luchter, R., Luna, A., Ma, J.,
Maheshwari, M., Mapua, P., Matlin, R., Martindale, A., Martinez, E.,
Massery, E., Mawhinney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,
Mosser, M., Neal, D., Mitchell, L., Mohabbat, K., Morgan, M., Morris, S.,
Nguyen, N., Nickerson, E., Niokekhwa, S., Ogun, M., Okunolu, G.,
Peters, L., Pickens, R., Preece, A., Payton, B., Peery, J., Perez, L.,
Rivers, M., Rojks, A., Rojudo, R., Ruffin, E., Pu, L.L., Quiles, M., Ren, Y.,
Schcherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
Sodergren, E., Somalvi, T., Sparks, A., Stanley, H., Stone, H.,
Tanton, A., Svatler, A., Taber, P., Tamestis, K., Tamestis, K., Tang, H.,
Tang, S., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, Q.,
Tsimak, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q.,
Williams, G., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Wu, C., Wu, Y.F., Williamson, A., Wleciyk, R., Wooden, S., Worley, K.,
Zack, G., Wu, Y.F., Zhou, J., Zorrilla, S., Zorrilla, S., Nelson, D.,

TITLE
JOURNAL
REFERENCE
AUTHORS

(bases 1 to 123770)

TITLE
JOURNAL

Direct Submission
Submitted (09-MAY-2002)

REFERENCE

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
(bases 1 to 123770)

TITLE
JOURNAL

Submitted (18-JUL-2003) Human Genome

COMMENT

Jul 14, 2002 this sequence was deposited in the GenBank database under accession number AF141414.1. The authors thank Dr. Robert H. Miller for his critical reading of the manuscript and Dr. Robert H. Miller for his critical reading of the manuscript.

Center: Baylor College of Medicine
Center code: BCM

Contact: hgsc-help@bcm.tmc.edu

Center clone name: CH230-330L5

Summary Statistics

Chemistry: Dye-terminator Big Dye. 100% of read

Assembly program: Phrap; version 0.990329

Consensus quality: 99.583 bases at least 1000

Consensus quality: 101766 bases at least Q20

NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

NOTE: This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1039	1038: contig of 1038 bp in length
1039	1138: gap of unknown length
1139	2213: contig of 1075 bp in length
2214	2313: gap of unknown length
2314	3496: contig of 1183 bp in length
3497	3597: gap of unknown length
3597	4862: contig of 1366 bp in length
4862	5062: gap of unknown length
5063	6379: contig of 1317 bp in length
6380	6479: gap of unknown length
6480	8102: contig of 1623 bp in length
8103	8202: gap of unknown length
8203	9610: contig of 1308 bp in length
9611	9610: gap of unknown length
9611	10987: contig of 1377 bp in length
10988	11087: gap of unknown length
11088	12208: contig of 1121 bp in length
12209	12308: gap of unknown length
12309	13660: contig of 1352 bp in length
13661	13760: gap of unknown length
13761	15627: contig of 1866 bp in length
15627	15626: gap of unknown length
15627	17972: contig of 2246 bp in length
17973	18072: gap of unknown length
18073	20240: contig of 2168 bp in length
20241	23269: contig of 2929 bp in length
20341	23369: gap of unknown length
23270	25441: contig of 2072 bp in length
23370	25542: gap of unknown length
25442	25542: contig of 3538 bp in length
25542	29080: gap of unknown length
29080	29179: gap of unknown length
29180	33526: contig of 4347 bp in length
33527	33626: gap of unknown length
33627	37047: contig of 3421 bp in length
37048	37147: gap of unknown length
37148	40494: contig of 3347 bp in length
37149	40594: gap of unknown length
40595	44084: contig of 3490 bp in length
44085	44184: gap of unknown length
44185	44306: contig of 4122 bp in length
44307	48406: gap of unknown length
48407	53058: contig of 4652 bp in length
53059	53158: gap of unknown length
53159	60471: contig of 7312 bp in length
60471	60570: gap of unknown length
60571	67470: contig of 6900 bp in length
67471	72023: gap of unknown length
72024	72123: contig of 4453 bp in length
72125	82882: gap of unknown length
82883	82882: contig of 10759 bp in length
82883	82982: gap of unknown length
82983	88103: contig of 6123 bp in length
88106	88205: gap of unknown length
89106	98084: contig of 8879 bp in length

21718 23427: contig of 1710 bp in length
 23428 23527: gap of unknown length
 23528 25577: contig of 2050 bp in length
 25578 25677: gap of unknown length
 25678 26828: contig of 1151 bp in length
 26829 26828: gap of unknown length
 26928 28047: contig of 1119 bp in length
 28048 28147: gap of unknown length
 28148 29246: contig of 1099 bp in length
 29247 29347: gap of unknown length
 29348 30841: contig of 1495 bp in length
 30842 30941: gap of unknown length
 30942 32355: contig of 1314 bp in length
 32356 32355: gap of unknown length
 33373 33472: contig of 1017 bp in length
 33473 35208: contig of 1736 bp in length
 35209 35308: gap of unknown length
 35309 36753: contig of 1445 bp in length
 36754 36853: gap of unknown length
 36854 38391: contig of 1538 bp in length
 38392 38491: gap of unknown length
 38492 40463: contig of 1972 bp in length
 40464 40563: gap of unknown length
 40564 42523: contig of 1960 bp in length
 42524 42523: gap of unknown length
 42624 43850: contig of 1227 bp in length
 42624 43851: gap of unknown length
 43851 43950: gap of unknown length
 43951 45002: contig of 1052 bp in length
 45003 45102: gap of unknown length
 45103 46452: contig of 1350 bp in length
 46453 46552: gap of unknown length
 46553 47971: contig of 1419 bp in length
 47972 48071: gap of unknown length
 48072 50174: contig of 2103 bp in length
 50175 50274: gap of unknown length
 50375 52391: contig of 2117 bp in length
 52392 52491: gap of unknown length
 52492 53721: contig of 1230 bp in length
 53722 53821: gap of unknown length
 53822 55732: contig of 1911 bp in length
 55733 55832: gap of unknown length
 55833 57430: contig of 1598 bp in length
 57431 57531: gap of unknown length
 57531 59819: contig of 2289 bp in length
 59820 59919: gap of unknown length
 59920 61778: contig of 1859 bp in length
 61779 61878: gap of unknown length
 61879 63710: contig of 1831 bp in length
 63710 63809: gap of unknown length
 63810 66227: contig of 2418 bp in length
 66228 66327: gap of unknown length
 66328 67583: contig of 1256 bp in length
 67584 67683: gap of unknown length
 67684 68700: contig of 1017 bp in length
 68701 68800: gap of unknown length
 68801 70067: contig of 1267 bp in length
 70068 70167: gap of unknown length
 70168 72507: contig of 2340 bp in length
 72508 72607: gap of unknown length
 72608 74884: contig of 2277 bp in length
 74885 74984: gap of unknown length
 74985 77133: contig of 2149 bp in length
 77134 77233: gap of unknown length
 77234 80004: contig of 2771 bp in length
 80005 80104: gap of unknown length
 82540 82540: contig of 2436 bp in length
 82541 82640: gap of unknown length
 82641 84674: contig of 2034 bp in length
 84675 84774: gap of unknown length
 84775 86258: contig of 1484 bp in length
 86259 86358: gap of unknown length
 86359 88249: contig of 1891 bp in length

Query Match 67.9%; Score 19; DB 2; Length 153987;
 Best Local Similarity 81.5%; Pred. No. 26+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 GTTCGGGATGTCGTCATGCTGT 28
 Db 77771 GTTCGCGGATGTCCTTCCTGCTGT 77745

RESULT 13
 AC117647/c
 LOCUS
 DEFINITION
 AC117647
 VERSION
 AC117647.2 GI:22123324
 KEYWORDS
 HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 house mouse.
 Mus musculus
 house mouse.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 AUTHORS
 1 (bases 1 to 179572)
 Birren, B., Nussbaum, C., and Lander, E.
 Mus musculus, clone RP23-244W7
 Unpublished
 2 (bases 1 to 179572)
 Birren, B., Linton, L., Nussbaum, C., Lander, E., Allen, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., Deatrellano, K., Dewar, R., Diaz, J.S., Dodge, S.,
 Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand, P., Jones, C.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R.,
 Landers, T., Lehocsky, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, R., Meldrum, J., Meneus, L.,
 Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, M., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (10-Apr-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 179572)
 Birren, B., Nussbaum, C., Lander, E., Allen, A., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., Deatrellano, K., Dewar, R., Diaz, J.S., Dodge, S.,
 Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand, P., Jones, C., Kamat, A.,
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kells, C.,
 Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J.,
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, R., Meldrum, J.,
 Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, M., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 3 (bases 1 to 179572)
 Birren, B., Nussbaum, C., Lander, E., Allen, A., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., Deatrellano, K., Dewar, R., Diaz, J.S., Dodge, S.,
 Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand, P., Jones, C., Kamat, A.,
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kells, C.,
 Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J.,
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, R., Meldrum, J.,
 Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, M., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced gi:20128406.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L23595

Center clone name: 244.M.7

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 172421 bases at least Q40
Consensus quality: 175743 bases at least Q30
Consensus quality: 176806 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 177472; sum-of-ctrls
Quality coverage: 5.0 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

1      832 931: gap of 100 bp
      932 2184: contig of 1253 bp in length
      2185 2284: gap of 100 bp
      2285 3203: contig of 921 bp in length
      3206 3305: gap of 100 bp
      3306 4649: contig of 1344 bp in length
      4650 4749: gap of 100 bp
      4750 7578: contig of 2829 bp in length
      7579 7678: gap of 100 bp
      7679 10016: contig of 2338 bp in length
      10017 10116: gap of 100 bp
      10117 11576: contig of 1460 bp in length
      11577 11676: gap of 100 bp
      11677 14279: contig of 2603 bp in length
      14280 14379: gap of 100 bp
      14380 52253: contig of 37874 bp in length
      52254 52353: gap of 100 bp
      52354 56294: contig of 3941 bp in length
      56295 56394: gap of 100 bp
      56395 62939: contig of 6545 bp in length
      62940 63039: gap of 100 bp
      63040 67332: contig of 4293 bp in length
      67333 67432: gap of 100 bp
      67433 74902: contig of 7470 bp in length
      74903 75002: gap of 100 bp
      75003 85025: contig of 10023 bp in length
      85026 85125: gap of 100 bp
      85126 91380: contig of 6255 bp in length
      91381 91480: gap of 100 bp
      91481 98876: contig of 7396 bp in length
      98877 98976: gap of 100 bp
      98977 106865: contig of 7889 bp in length
      106866 106965: gap of 100 bp
      106966 119527: contig of 12562 bp in length
      119528 119627: gap of 100 bp
      119628 135356: contig of 15729 bp in length

```

```

* 135357 135456: gap of 100 bp
* 135457 152890: contig of 17434 bp in length
* 152891 152990: gap of 100 bp
* 152991 176149: contig of 22159 bp in length
* 176150 176249: gap of 100 bp
* 176250 179572: contig of 3323 bp in length.
  location/Qualifiers
1. 179572
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /clone="RP23-244M7"
  /clone_1lb="RPC1-23 Female Mouse BAC"
1. 831
  /note="assembly-fragment"
  /clone_end:SP6
  /vector_side:left"

```

```

misc_feature
932..2184
/note="assembly-fragment"
2285..3205
/note="assembly-fragment"
3306..4649
/note="assembly-fragment"
4750..7578
/note="assembly-fragment"
7679..10016
/note="assembly-fragment"
10117..11576
/note="assembly-fragment"
11677..14279
/note="assembly-fragment"
14380..52253
/note="assembly-fragment"
52354..56294
/note="assembly-fragment"
56395..62939
/note="assembly-fragment"
63040..67332
/note="assembly-fragment"
67433..74902
/note="assembly-fragment"
75003..85025
/note="assembly-fragment"
85126..91380
/note="assembly-fragment"
91481..98876
/note="assembly-fragment"
98977..106865
/note="assembly-fragment"
106966..119527
/note="assembly-fragment"
119628..135356
/note="assembly-fragment"
135457..152890
/note="assembly-fragment"
152991..176149
/note="assembly-fragment"
176250..179572
/note="assembly-fragment"
  /clone_end:r7
  /vector_side:right"

```

BASE COUNT 55686 a 32751 c 33323 g 55711 t 2101 others

Query Match 67.9% Score 19; DB 2; Length 179572;
Best Local Similarity 81.5%; Pred. NO. 2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GTTGTGGATAGCCGTCATGCGT 28
||||| ||||| ||||| ||||| |||||
DB 158787 GTTCTTGAAGTAAGTCATGCTT 158761

RESULT 14

AC009921
LOCUS 189086 bp DNA linear HTG 26-MAR-2000
DEFINITION Homo sapiens clone RP11-115018, WORKING DRAFT SEQUENCE, 16
AC009921
AC009921
AC009921.4 GI:7329301
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens, clone RP11-115018
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 189086).
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baker, J., Baldwin, J., Barna, N., Beckert, R., Benn, J., Brown, A.,
Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeRubeis, K., Depayre, E., Devon, K., Dewar, K.,
Donelan, L., Doyle, M., Ferrel, P., Fitzhugh, W., Forrest, C.,
Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
Karatas, A., Lehoczy, J., Lieu, C., Locke, K., MacDonald, P.,
Marquis, N., McEwan, P., McGuck, A., McKenna, K., McLaughlin, J.,
Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
Stevage, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
Direct Submission
Submitted (08-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6479138.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L1914
Center clone name: 115.O.18
Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179818 bases at least Q40
Consensus quality: 183946 bases at least Q40
Consensus quality: 185564 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 187586; sum-of-coverage
Quality coverage: 5.2 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-coverage
NOTE: This is a 'working draft' sequence. It currently
consists of 16 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1064::contig of 1064 bp in length
1065 1164::gap of 100 bp
1165 2211::contig of 1047 bp in length
2212 2311::gap of 100 bp
2312 3615::contig of 1304 bp in length
3616 3715::gap of 100 bp
3716 5319::contig of 1604 bp in length

FEATURES
Source
* 5320 5419::gap of 100 bp
* 5420 7668::contig of 2249 bp in length
* 7669 7768::gap of 100 bp
* 7769 12506::contig of 4738 bp in length
* 12507 12606::gap of 100 bp
* 12607 17063::contig of 4457 bp in length
* 17064 17163::gap of 100 bp
* 17164 24729::contig of 7566 bp in length
* 24730 24829::gap of 100 bp
* 24830 35834::contig of 11005 bp in length
* 35835 35934::gap of 100 bp
* 35935 46466::contig of 10532 bp in length
* 46467 46566::gap of 100 bp
* 46567 61624::contig of 15058 bp in length
* 61625 61724::gap of 100 bp
* 61725 77509::contig of 15785 bp in length
* 77510 77609::gap of 100 bp
* 77610 94997::contig of 17388 bp in length
* 94998 95097::gap of 100 bp
* 95098 120488::contig of 25391 bp in length
* 120489 120589::gap of 100 bp
* 120589 147913::contig of 27325 bp in length
* 147914 148013::gap of 100 bp
* 148014 189086::contig of 41073 bp in length.
Location/Qualifiers
1. 189086
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-115018"
/clone_1bp="RPCT-11 Human Male BAC"
1. 1064
/note="assembly-fragment"
1165. 2211
/note="assembly-fragment"
2312. 3615
/note="assembly-fragment"
3716. 5319
/note="assembly-fragment"
5420. 7668
/note="assembly-fragment"
7769. 12506
/note="assembly-fragment"
12607. 17063
/note="assembly-fragment"
clone_end:77
vector_side:right
17164. 24729
/note="assembly-fragment"
24830. 35834
/note="assembly-fragment"
35935. 46466
/note="assembly-fragment"
46567. 61624
/note="assembly-fragment"
clone_end:SP6
vector_side:right
61725. 77509
/note="assembly-fragment"
77610. 94997
/note="assembly-fragment"
95098. 120488
/note="assembly-fragment"
120589. 147913
/note="assembly-fragment"
148014. 189086
/note="assembly-fragment"

BASE COUNT 61951 a 33004 c 33086 g 59542 t 1503 others
ORIGIN

Query Match 67.9%; Score 19; DB 2; Length 189086;
Best Local Similarity 81.5%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 GTTCGTGGATAGTCGTCATGTCGT 28
 Db 88210 GTTCCTGGTATGTCATGTCGT 88236
 RESULT 15
 AL354896/C
 LOCUS
 DEFINITION Human DNA sequence from clone RP11-512M17 on chromosome 13,
 complete sequence.
 ACCESSION AL354896 209743 bp DNA linear PRI 29-NOV-2000
 VERSION AL354896
 KEYWORDS AL354896.16 GI:11322813
 SOURCE HTG.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 209743)
 AUTHORS Lovell, J.
 TITLE Direct Submission
 JOURNAL Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 Requests: clonerequests@sanger.ac.uk
 On Nov 23, 2000 this sequence version replaced gi:113221907.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated repeat sequence elements. Where the sequence is
 ambiguous, there is an annotation using the 'unsube' feature key.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone configs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr13
 RP11-512M17 is from the library RPCR-11.2 constructed at the
 Roswell Park Cancer Institute by the group of Pieter de Jong. For
 further details see http://bacpac.med.buffalo.edu/
 VECTOR: pBAC3.6
 This sequence is the entire insert of clone RP11-512M17 The true
 left end of clone RP11-545P6 is at 121889 in this sequence. The
 true right end of clone RP11-315A9 is at 159267 in this sequence.
 FEATURES
 Source
 Location/Qualifiers
 1..209743
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone="RP11-512M17"
 /clone_lib="RPCR-11.2"
 1..265
 /note="MST2 repeat: matches 160. 426 of consensus"
 repeat_region 270..1579
 /note="L1P12 repeat: matches -55. 1238 of consensus"
 repeat_region 1572..2157
 /note="L1P12 repeat: matches -1416. -839 of consensus"
 repeat_region 2153..4016
 /note="L1 repeat: matches 2124. 4006 of consensus"
 misc_feature 2830..2867
 /note="Sequence from overlapping clone
 BA315A9(AL162498). Assembly confirmed by restriction
 digest."
 repeat_region 3946..4064
 /note="L1M2 repeat: matches 2455. 2571 of consensus"
 repeat_region 4095..4188

repeat_region /note="47 copies 2 mer ct 69% conserved"
 4192..4433
 /note="AluSc repeat: matches 135. 309 of consensus"
 repeat_region 4437..6353
 /note="L1M2 repeat: matches 132. 2403 of consensus"
 repeat_region 6521..6942
 /note="L1M2 repeat: matches -707. -278 of consensus"
 repeat_region 7203..9020
 /note="L1P repeat: matches 2297. 4115 of consensus"
 repeat_region 10243..11214
 /note="L1M1 repeat: matches 1555. 2532 of consensus"
 repeat_region 11210..11568
 /note="L1P17 repeat: matches 5749. 6118 of consensus"
 repeat_region 11657..11957
 /note="AluY repeat: matches 1. 299 of consensus"
 repeat_region 11969..12276
 /note="L1M3 repeat: matches -870. -580 of consensus"
 repeat_region 12545..13175
 /note="L2 repeat: matches 2114. 2710 of consensus"
 repeat_region 13304..13609
 /note="AluX repeat: matches 1. 312 of consensus"
 repeat_region 13748..13798
 /note="L17 copies 3 mer ttg 80% conserved"
 repeat_region 15340..15435
 /note="M1711 repeat: matches 318. 410 of consensus"
 repeat_region 16287..16584
 /note="AluSg repeat: matches 1. 306 of consensus"
 repeat_region 16744..17997
 /note="Tigge3b repeat: matches 4. 1224 of consensus"
 repeat_region 18397..18771
 /note="THEIC repeat: matches 1. 371 of consensus"
 repeat_region 18772..20391
 /note="THEIC-internal repeat: matches 1. 1578 of
 consensus"
 repeat_region 20394..20762
 /note="THEIC repeat: matches 1. 360 of consensus"
 repeat_region 24284..24331
 /note="L16 copies 3 mer agy 75% conserved"
 misc_feature 25585..25861
 /note="Sequence from overlapping clone
 BA315A9(AL162498). Assembly confirmed by restriction
 digest."
 repeat_region 27691..27854
 /note="MST2 repeat: matches 1. 177 of consensus"
 repeat_region 27854..27960
 /note="THEIC repeat: matches 316. 422 of consensus"
 misc_feature 28176..28487
 /note="Sequence from overlapping clone
 BA315A9(AL162498). Assembly confirmed by restriction
 digest."
 repeat_region 28418..28850
 /note="L1M2 repeat: matches 5318. 5765 of consensus"
 repeat_region 28883..28924
 /note="L21 copies 2 mer ta 76% conserved"
 repeat_region 28945..29194
 /note="L1M2 repeat: matches 4978. 5229 of consensus"
 repeat_region 29235..29844
 /note="L1M2 repeat: matches 4297. 4914 of consensus"
 repeat_region 30076..30415
 /note="MER2 repeat: matches 2. 344 of consensus"
 misc_feature 30525..30717
 /note="Sequence from overlapping clone
 BA315A9(AL162498). Assembly confirmed by restriction
 digest."
 repeat_region 35507..35598
 /note="AluIo/FRAM repeat: matches 157. 249 of consensus"
 repeat_region 35711..35961
 /note="L1M3 repeat: matches 6054. 6304 of consensus"
 repeat_region 36289..36663
 /note="L1M4 repeat: matches 5911. 6299 of consensus"
 repeat_region 37109..37424
 /note="AluSg repeat: matches 3. 308 of consensus"
 repeat_region 38353..39269

```

repeat_region /note="L1MA10 repeat: matches 5409. .6317 of consensus"
39487. .39579
/note="MIR repeat: matches 166. .261 of consensus"
39604. .39737
/note="L1ME repeat: matches 5308. .5443 of consensus"
40178. .40217
/note="20 copies 2 mer ct 97% conserved"
40226. .40654
/note="L1M4 repeat: matches 4141. .4604 of consensus"
42776. .42883
/note="54 copies 2 mer aa 63% conserved"
42977. .43059
/note="L2 repeat: matches 2612. .2691 of consensus"
43186. .43309
/note="MLTIE repeat: matches 296. .419 of consensus"
43887. .44226
/note="L2 repeat: matches 1882. .2236 of consensus"
44749. .44859
/note="L1MC3 repeat: matches 6701. .6810 of consensus"
44903. .44935
/note="L1MC4 repeat: matches 6868. .6900 of consensus"
45532. .45978
/note="L1R40b repeat: matches 4. .462 of consensus"
46667. .46847
/note="Sequence from overlapping clone
ba315A9(Alu62458). Assembly confirmed by restriction
digest."
46980. .47254
/note="HERVLA40 repeat: matches 1. .278 of consensus"
48030. .48134
/note="L2 repeat: matches 2015. .2117 of consensus"
48140. .48519
/note="THEIC repeat: matches 1. .369 of consensus"
48530. .48581
/note="26 copies 2 mer aa 71% conserved"
48650. .48700
/note="L2 repeat: matches 2211. .2261 of consensus"
51749. .51964
/note="L2 repeat: matches 1825. .2054 of consensus"
52481. .53602
/note="L1R28 repeat: matches 1. .1735 of consensus"
54284. .54627
/note="L1PA12 repeat: matches 5746. .6108 of consensus"
54630. .55318
/note="L1 repeat: matches 4538. .5260 of consensus"
56627. .57090
/note="MLTID repeat: matches 9. .505 of consensus"
57545. .58643
/note="L1MA8 repeat: matches 5148. .6287 of consensus"
58669. .58845
/note="THEIC repeat: matches 1. .177 of consensus"
58846. .59077
/note="L1PA5 repeat: matches 5909. .6143 of consensus"
59078. .59332
/note="THEIC repeat: matches 177. .425 of consensus"
59333. .59676
/note="L1MA8 repeat: matches 4787. .5125 of consensus"
59755. .59837
/note="L1MD1 repeat: matches 6133. .6216 of consensus"
59829. .60433
/note="L1MD repeat: matches 83. .708 of consensus"

Query Match 67.9%; Score 19; DB 9; Length 209743;
Best Local Similarity 81.5%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 2 GTTCGGGATAGTCGTCATGCTGT 28
Db 181468 GTTCCTGGGATCTGTCATGCTGT 181442

```

Search completed: July 10, 2003, 19:20:39
 Job time : 251.625 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 59.0831 Seconds
(without alignments)
1067.241 Million cell updates/sec

Title: US-10-087-631B-9

Perfect score: 28
Sequence: 1 cgttcgtaggtagctcgtcgtgtgtt 28

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	65.7	510	24	ABO41020
2	18.4	65.7	510	24	ABO41021
3	18.4	65.7	565	24	ABO19830
4	18.4	65.7	565	24	ABO19831
5	18.4	65.7	2636	23	AB104197
6	18.4	65.7	7330	23	AB104196
7	18.4	64.3	6265	14	AAQ49460
8	17.6	62.9	34088	23	AA559566
9	17.4	62.1	297	24	ABN76849

C	10	17.4	62.1	346	21	ABO62757	Mycobacterium tube
C	11	17.4	62.1	541	24	ABO46568	Oligonucleotide fo
C	12	17.4	62.1	541	24	ABO46569	Oligonucleotide fo
C	13	17.4	62.1	654	24	ABN24612	Human ORF polynuc
C	14	17.4	62.1	1149	22	AAE65378	Drosophila gustato
C	15	17.4	62.1	1755	23	ABL27479	Drosophila melanog
C	16	17.4	62.1	2976	22	AAF81351	Quorum sensing con
C	17	17.4	62.1	3836	15	AAO53997	Vitamin D hydroxyl
C	18	17.4	62.1	4337	23	ABL27478	Drosophila melanog
C	19	17.4	62.1	5356	21	AAA46169	GST-GFP fusion pro
C	20	17.4	62.1	6494	21	AAA46170	GFP-Sml-7 fusion p
C	21	17.4	62.1	8549	18	AAV74316	Staphylococcus aur
C	22	17.4	62.1	11460	21	AAZ87211	VEE replicon compr
C	23	17.4	62.1	11795	22	AAH98981	Human EST-derived
C	24	17.4	62.1	11835	23	AA584466	DNA encoding novel
C	25	17.4	62.1	43095	21	AAA68254	Bacteriophage 3A c
C	26	17.4	62.1	4403765	22	AA199683	Mycobacterium tube
C	27	17.4	62.1	4411529	22	AA199682	DNA encoding novel
C	28	17.2	61.4	462	23	AA580504	Gene #3877 used to
C	29	17.2	61.4	1247	24	ABN97379	Human polynucleoti
C	30	17.2	61.4	1310	22	AA159616	Human polynucleoti
C	31	17.2	61.4	1320	22	AA157830	Human gene signatu
C	32	17.2	60.7	135	16	AA193448	Human secreted pro
C	33	17.2	60.7	135	21	AAAC07458	Human purH amplico
C	34	17.2	60.7	467	21	AAA28168	Drosophila melanog
C	35	17.2	60.7	580	23	ABL11249	DNA encoding novel
C	36	17.2	60.7	588	23	AA570525	Fennugreek galactom
C	37	17.2	60.7	725	21	AA255634	Fennugreek galactom
C	38	17.2	60.7	739	24	ABK68668	Human DNA for olfa
C	39	17.2	60.7	939	24	ABK68670	Human DNA for olfa
C	40	17.2	60.7	940	24	ABO40952	Oligonucleotide fo
C	41	17.2	60.7	940	24	ABO40953	Oligonucleotide fo
C	42	17.2	60.7	1290	19	AAV53208	Human olfactory OL
C	43	17.2	60.7	1314	21	AA255631	Fennugreek galactom
C	44	17.2	60.7	1415	21	AA255635	Fennugreek galactom
C	45	17.2	60.7	2696	14	AAO41061	Ap Serotype 7 60KD

ALIGNMENTS

RESULT 1	ABO41020	standard; DNA; 510 BP.
ID	ABO41020	
AC	ABO41020	
XX		
DT	12-JUL-2002 (first entry)	
XX		
DE	Oligonucleotide for detecting cytosine methylation seq ID NO 27611.	
XX		
KW	Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;	
KW	drug; side effect; cancer; central nervous system; cardiovascular;	
KW	gastrointestinal; respiratory system; single nucleotide polymorphism;	
KW	SNP; cell differentiation; ds.	
OS	Homo sapiens.	
XX		
PN	WO200218632-A2.	
XX		
PD	07-MAR-2002.	
XX		
PF	01-SEP-2001; 2001WO-EP10074.	
XX		
PR	01-SEP-2000; 2000DE-1043826.	
XX		
PR	05-SEP-2000; 2000DE-1044543.	
XX		
PA	(EPIC-) EPICENOMICS AG.	
XX		
PI	Olek A, Piepenbrock C, Berlin K, Guetig D;	
XX		
DR	WPI; 2002-371829/40.	
XX		

PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PS amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (1) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (11) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

SO Sequence 510 BP; 112 A; 45 G; 151 C; 202 T; 0 other;

Query Match 65.7%; Score 18.4; DB 24; Length 510;
 Best Local Similarity 78.6%; Pred. No. 43;
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CGTTCGTGGATGATCCGTATGCTGT 28
 DB 477 CGTTCGTGGATGATGTTAGTGTCTT 504

RESULT 2

ABQ1021/C
 ID ABQ1021 standard; DNA; 510 BP.
 AC ABQ1021;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 27612.

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PS amplicons from chemically treated DNA

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (1) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (11) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

SO Sequence 510 BP; 202 A; 151 C; 45 G; 112 T; 0 other;

Query Match 65.7%; Score 18.4; DB 24; Length 510;
 Best Local Similarity 78.6%; Pred. No. 43;
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CGTTCGTGGATGATCCGTATGCTGT 28
 DB 34 CGTTCGTGGATGATGTTAGTGTCTT 7

RESULT 3

ABQ19830
 ID ABQ19830 standard; DNA; 565 BP.
 AC ABQ19830;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 6421.

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful.
 PT for diagnosis and prognosis, comprises selective hybridization of
 PS amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABO13410-ABO54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX Sequence 565 BP; 123 A; 49 C; 167 G; 226 T; 0 other;

Query Match 65.7%; Score 18.4; DB 24; Length 565;

Best Local Similarity 78.6%; Pred. No. 44;

Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CGTGTGGGATAGTCGTCATGCTGT 28

DB 376 CGTGTGGGATAGTCGTCATGCTGT 403

RESULT 4

ID ABO19831/C standard; DNA; 565 BP.

XX ABO19831;

XX 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 6422.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

XX drug; side effect; cancer; central nervous system; cardiovascular;

XX gastrointestinal; respiratory system; single nucleotide polymorphism;

XX SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIDENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI: 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful

XX for diagnosis and prognosis, comprises selective hybridization of

XX amplicons from chemically treated DNA

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a

XX genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABO13410-ABO54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX Sequence 565 BP; 226 A; 167 G; 49 C; 123 T; 0 other;

Query Match 65.7%; Score 18.4; DB 24; Length 565;

Best Local Similarity 78.6%; Pred. No. 44;

Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CGTGTGGGATAGTCGTCATGCTGT 28

DB 190 CGTGTGGGATAGTCGTCATGCTGT 163

RESULT 5

ID ABL04197/C standard; cDNA; 2636 BP.

XX ABL04197;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 7073.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EM;

XX WPI: 2001-656860/75.

XX P-PSDB; ABB60094.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions

XX Claim 1; SEQ ID NO 7073; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins

SQ Sequence 2636 BP; 810 A; 665 C; 649 G; 512 T; 0 other.

```

22; conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

[illegible]

AC ABL04196;

XX

Prasopbila: developmental biology

XXXXXX

XX

XX
C
T
T
Z
Z

XXXXXXXXXXXX

23-MAR-2001; 2001WO-US09231

TCZC050 CNT007 / T007

DATE	DESCRIPTION	AMOUNT	BALANCE
1950			
1951			
1952			
1953			
1954			
1955			
1956			
1957			
1958			
1959			
1960			
1961			
1962			
1963			
1964			
1965			
1966			
1967			
1968			
1969			
1970			
1971			
1972			
1973			
1974			
1975			
1976			
1977			
1978			
1979			
1980			
1981			
1982			
1983			
1984			
1985			
1986			
1987			
1988			
1989			
1990			
1991			
1992			
1993			
1994			
1995			
1996			
1997			
1998			
1999			
2000			
2001			
2002			
2003			
2004			
2005			
2006			
2007			
2008			
2009			
2010			
2011			
2012			
2013			
2014			
2015			
2016			
2017			
2018			
2019			
2020			
2021			
2022			
2023			
2024			
2025			
2026			
2027			
2028			
2029			
2030			
2031			
2032			
2033			
2034			
2035			
2036			
2037			
2038			
2039			
2040			
2041			
2042			
2043			
2044			
2045			
2046			
2047			
2048			
2049			
2050			
2051			
2052			
2053			
2054			
2055			
2056			
2057			
2058			
2059			
2060			
2061			
2062			
2063			
2064			
2065			
2066			
2067			
2068			
2069			
2070			
2071			
2072			
2073			
2074			
2075			
2076			

XXX

WPI: 2001-656860/75.

[illegible]

interactions -

...the importance of...

(ABB57737-ABB72072).

[illegible]

Query Match	Count	10.1	22.2
65.7%			

Statistics	Conservative	Mismatches	Indels
42;	10;	6;	0;

[illegible]

1. *Chlorophyll a* (Chl *a*)

[illegible]

XX
XX

XX.
XX.
Trea amido]raco. ndr

aa
05 *Saccharomyces cerevisiae*.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

```
/*tag= a
```

XX

[illegible][illegible]

XX PA (TOYM) TOYOBQ KR

DR · WPI; 1993-338925/43

XX

... comprising the DNA

xx. page 10 1/1 1/pp; japanese

[illegible]

Query Match	64 38	Score	%
		10	100

```

##### 22, conservative 0; mismatches 5; indels 0; gaps 0;

```

00000 CCGTCTGAGCCGGATGATG 3/08

D AAS59566 standard; DNA; 34088 BP

X

Preferential treatment

SAPHO syndrome: synovitis, acne, pustulosis, pyoderma, osteitis

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

100

PN WO200181581-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX 02-JUN-2000; 2000US-208841P.
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX WPI; 2001-616774/71.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris -
XX
XX Claim 1; SEQ ID No 61; 1069pp; English.
XX
XX Sequences AAS59506-AAS59804 represent DNA molecules encoding
XX CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
XX CC associated DNA sequences are used in the treatment, prevention and
XX CC diagnosis of medical conditions caused by P. acnes. The disorders include
XX CC SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and
XX CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
XX CC in infections of bone, joints and the central nervous system, however it
XX CC is particularly involved in the inflammatory lesions associated with acne
XX CC vulgaris. A method for detecting the presence or absence of P. acnes in a
XX CC patient comprises contacting a sample with a binding agent that binds to
XX CC the proteins of the invention and determining the amount of bound protein
XX CC in the sample. The polypeptides may be used as antigens in the production
XX CC of antibodies specific for P. acnes proteins. These antibodies can be
XX CC used to downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
XX CC polypeptides shown in AAV54556-AAV54928 and AAV67560-AAV67562.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 34088 BP; 6991 A; 11002 C; 9453 G; 6639 T; 3 other;
XX
XX Query Match 62.9%; Score 17.6; DB 23; Length 34088;
XX Best Local Similarity 83.3%; Pred. No. 1.9e+02;
XX Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX 1 CGTGTGGGATGATCGTCGTCATCG 24
XX DB 22579 CGTGTGGGATGATGACCGTCATCG 22556
XX
XX RESULT 9
XX ABN76849/C
XX ID ABN76849 standard; cDNA; 297 BP.
XX
XX ABN76849;
XX
XX 08-JUL-2002 (first entry).
XX
XX Human ORF1796 cDNA, SEQ ID NO:3591.
XX
XX Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
XX KM disease monitoring; cytokine; cell proliferation; cell differentiation;
XX KM immune modulation; haematopoiesis regulation; tissue growth;
XX KM angiogenesis; activin; inhibitor; chemotactic; chemokinetic; hemostatic;
XX KM thrombolytic; tumour inhibition; bodily characteristics; fertility;
XX KM behaviour; cancer; proliferative disorder; neurological disorder;
XX KM cardiovascular disease; immune system disorder; organ transplantation;

KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulvarey;
KW vasotropic; antipsoriatic; antidiabetic; cytosolic; neutrotic;
KW neuroprotective; antithrombotic; antithrombotic; thrombolytic;
KW cardiant; hypotensive; antihypertensive; antineoplastic; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200190366-A2.
XX
XX 29-NOV-2001.
XX
XX 24-MAY-2001; 2001WO-US17076.
XX
XX 24-MAY-2000; 2000US-206690P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Leach MD, Shinkels RA;
XX
XX WPI; 2002-106200/14.
XX DR P-PSDB; ABP32823.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX PT preventing and treating cardiovascular disease, neurodegenerative,
XX PT hyperproliferative disorders and disorders related to organ
XX PT transplantation -
XX
XX Claim 1; Page 1157; 2508pp; English.
XX
XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins
XX CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
XX CC ABN79587 represent cDNAs encoding them. The invention also encompasses
XX CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
XX CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
XX CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
XX CC polynucleotides, the recombinant production of ORFX proteins, antibodies
XX CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
XX CC polypeptides, methods of screening for modulators of ORFX expression or
XX CC activity, and methods of screening individuals for a predisposition to an
XX CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
XX CC range of biological activities, such as cytokine, cell proliferation,
XX CC cell differentiation, immune modulation, haematopoiesis regulation,
XX CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
XX CC chemokinetic activity, haemostatic activity, thrombolytic activity,
XX CC receptor/ligand, antineoplastic activity, tumour inhibition activity,
XX CC and anti-infective activity, and may also be involved in the determination
XX CC of bodily characteristics, fertility and behaviour. ORFX proteins,
XX CC nucleic acids and antibodies may be used in the treatment of cancers,
XX CC other proliferative disorders such as psoriasis and benign tumours,
XX CC neurological disorders such as epilepsy and Alzheimer's disease,
XX CC cardiovascular diseases, immune system disorders, disorders related to
XX CC organ transplantation, disorders of tissue growth and regeneration,
XX CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol
XX CC storage disease, and infectious diseases caused by viral, bacterial,
XX CC fungal and other pathogens. ORFX nucleic acids may also be used as a
XX CC source of primers and probes, in the detection of ORFX genomic sequences
XX CC or transcripts, in the identification and cloning of homologous
XX CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
XX CC nucleic acids may additionally be used to produce transgenic animals
XX CC which may be useful for studying the function and/or activity of ORFX
XX CC protein, and in drug screening. The ORFX proteins may also be used as
XX CC immunogens to generate specific antibodies, which are useful in the
XX CC diagnosis, treatment and monitoring of ORFX-associated diseases.
XX
XX Sequence 297 BP; 66 A; 99 C; 78 G; 54 T; 0 other;
XX
XX Query Match 62.1%; Score 17.4; DB 24; Length 297;
XX Best Local Similarity 77.8%; Pred. No. 1.2e+02;
XX Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX
XX 2 GTTCTGGGATGATCGTCGTCATCGT 28

RESULT 10	
ABQ62757/c	
ID	ABQ62757 standard; DNA; 346 BP
XX	

16-AUG-2002: (first entry)

XX Mycobacterium tuberculosis BAC vector clone RV240SP6

KW detection; BAC vector; bacterial artificial chromosome; tuberculosis, gene; ds.

05 Mycobacterium tuberculosis
XX

PN W09954487-A2

PD 28-OCT-1999

PF 16-APR-1999; 99WO-IB00740
XX

16-APR-1998; 98US-0060756

(INSP) INST. PASTEUR.

Buchrieser-Brosch R, Gordon S, Billault A, Cole S, Pfl

WFL; 2000-013262/01

detection of Mycobacteria and for combating tuberculosis -

Claim 23; Page 70; 161pp; English

of interest that is present or is expressed in a genome of a first mycobacterial strain and that is absent or altered in a genome of a second mycobacterial strain, which is different from the first strain vectors, which are preferably immobilised, can be used to detect mycobacterial nucleic acids (genomic DNA, cDNA or mRNA) in biological samples. The polynucleotides identified are useful as probes or primers for detecting a given mycobacterium of interest. By aligning the polynucleotides contained in the recombinant BAC vectors it is possible to physically map a polynucleotide of mycobacterial origin in a biological sample. The methods and vectors from the present invention are useful in providing information for combating tuberculosis. It is possible to compare genomes between different strains or species and their non-pathogenic strains or species counterparts. AB063229 to AB063228 and ABH81227 to ABH81230 represent sequences used in the exemplification of the present invention.

sequence 346 BP; 53 A; 96 C; 125 G; 70 T; 2 other;

Best Local Similarity	score 1/.4;	DB 21;	Length, 346;
94.78%;	Pred. No. 1	2e+02;	

Conservative	Mismatches	Indels	Gaps
0	0	0	0

253 GTTCGTCGGATAGTCCGTC 235

RESULT 11

ABQ46568 standard; DNA; 541 BP

ABQ46568

XX 12-JUL-2002 (first entry)
DT
XX

XX
XX

cytosine methylation SEQ ID NO 33159.

MN, drug; cytosine methylation; 5'-CPG-3'; unaccl; cytosine; diagnosis;
 NM, tumor; side effect; cancer; central nervous system; cardiovascular;
 KW, gastrointestinal; respiratory system; single nucleotide polymorphism
 XX, SNP; cell differentiation; ds.

OS Homo sapiens
XY

PN W02U0218632-A2
XX

PD 07-MAR-2002.
XX

01-SEP-2001; 2001WO-EPI0074
XX

PR 05-SEP-2000; 2000DE-1044543

PA (EPiG-) EPIGENOMICS AG

Olek A, Piepenbrock C, Berlin K, Guettig D
PI xy

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA "useful"

amplicons from chemically treated DNA -

PS Claim 12; 56pp + Sequence Listing; 56pp; German
XX

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the classes of oligomers. From the ratio of labels hybridised to the two classes (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABO131410-AB054121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 541 BP; 82 A; 59 C; 201 G; 197 T; 2 other;

Best Local Similarity	02.16;	score 17.4;	DB 24;	Length 541;
	77.88;	Pred. No. 1	3e+02;	

matches	21;	conservative	0;	Mismatches	6;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

Y 2 GTTCGTGGGATAGTCGCCGTCATGGTGT 28
|||||
b 450 GTTCGTGGGATCGGACGTCGGGTGT 476

RESULT 12

ABQ46569 standard; DNA; 541 BP.

ABQ46569;

12 JUL-2002 (first entry)

KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis; gene; ss.
 XX
 XX Homo sapiens..
 XX
 XX WO200192523-A2.
 XX
 XX 06-DEC-2001.
 XX
 XX 29-MAY-2001; 2001WO-US10836.
 XX
 XX 30-MAY-2000; 2000US-206132P.
 XX 29-AUG-2000; 2000US-228716P.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 XX Shimkels RA, Leach MD;
 XX
 XX WPI: 2002-106308/14.
 XX
 XX P-PSDB: ABP08860.
 XX
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 XX preventing and treating cardiovascular disease, neurodegenerative,
 XX hyperproliferative disorders and autoimmune disorders
 XX
 XX Disclosure; SEQ ID 17701; 1037Pp; English.
 XX
 XX The present invention describes substantially purified human proteins
 XX (referred to as open reading frame, ORFX, where x is 1-11491 (see Table 1
 XX in the specification)). ABRN5762 to ABRN27252 encode the human ORFX
 XX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 XX treating or preventing a pathology associated with an ORFX-associated
 XX disorder in humans, and in the manufacture of a medicament for treating a
 XX syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 XX sequences can be used in gene therapy. ORFX sequences can be used in the
 XX treatment of cancer. hyperproliferative disorders, cirrhosis of liver,
 XX psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 XX osteoarthritis, neurodegenerative disorders, disorders related to organ
 XX transplantation, cardiovascular diseases, diabetes mellitus, systemic
 XX lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 XX storage disease, various immune deficiencies and disorders, infectious
 XX diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 XX arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 XX disease and autoimmune inflammatory eye disease. ORFX proteins are also
 XX useful for treating burns, incisions, ulcers, for treating osteoporosis,
 XX bone degenerative disorders, or periodontal disease, and for gut
 XX protection or regeneration and treatment of lung or liver fibrosis,
 XX reperfusion injury in various tissues and conditions resulting from
 XX systemic cytokine damage.
 XX N.B. The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 654 BP; 90 A; 194 C; 180 G; 190 T; 0 other;
 XX
 XX Query Match 62.1%; Score-17.4; DB 24; Length 654;
 XX Best Local Similarity 77.8%; Pred. No. 1.3e+02;
 XX Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 XX
 XX 1 CGTTCGTGGGATAGTCGCTCATGGTGT 27
 XX ||||||| || ||||| || |||||
 XX Db 344 CGTTCGTGATATGCTCCGCCAGGCTCT 370
 XX
 XX RESULT 14
 XX AAF63738
 XX ID AAF63738 standard; DNA; 1149 BP.

AC AAF63738;
 XX
 DT 03-APR-2001 (first entry)
 XX
 DE Drosophila gustatory receptor GR23A.1a DNA sequence.
 XX
 KM Gustatory receptor; fruit fly; taste; pheromone; semiochemical;
 XX crop damage; pest control; ds.
 OS
 XX Drosophila melanogaster.
 XX
 PN WO200077208-A2.
 XX
 PD 21-DEC-2000.
 XX
 PF 14-JUN-2000; 2000WO-US16211.
 XX
 PR 14-JUN-1999; 99US-0138668.
 XX
 PR 10-FEB-2000; 2000US-0181704.
 XX
 PA (UYVA) UNIV YALE.
 XX
 PI Carlson PJ, Clyne PJ, Mair CG;
 XX
 DR WPI; 2001-061873/07.
 XX
 DR P-PSDB; AAB75199.
 XX
 PT New isolated nucleic acid molecule encoding Drosophila gustatory
 PT Receptor protein useful for e.g. identification of compounds which may
 PT be used for pest management -
 XX
 PS Claim 3; Page 95-97; 227pp. English.
 XX
 CC This invention relates to polynucleotide sequences AAF63732 - AAF63777
 CC which encode Drosophila gustatory receptor proteins represented by
 CC sequences AAB75193 - AAB75238. The invention includes methods for
 CC determining gustatory receptor ligands. Also included is a method for
 CC modulating the expression of the DNA encoding the receptors. The DNA and
 CC protein sequences may be used for the identification of compounds,
 CC e.g. pheromones and other semiochemicals, which may be used for pest
 CC management. The DNA sequences may also be used for behavioural studies
 CC involving gustatory systems in various organisms. Also, the DNA sequences
 CC may also be used to track down gustatory receptor genes in insects that
 CC damage crops or transmit diseases.
 XX
 SQ Sequence 1149 BP; 255 A; 276 C; 292 G; 326 T; 0 other;
 Query Match 62.1%; Score 17.4; DB 22; Length 1149;
 Best Local Similarity 77.8%; Pred. NO. 1.4e+02;
 Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 OY 2 GTTCGTGGAGATGCGTCGATGCTT 28
 ||||||||| ||||| ||| ||
 Db 258 GTTCGTGGAGAACTCGTCTGGGATT 284
 RESULT 15
 ABL27479
 ID ABL27479 standard; DNA: 1755 BP.
 XX
 AC ABL27479;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 33910.
 XX
 KM Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX

PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 XX
 PF 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 33910; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (AAB57737-ABR20722).
 CC
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1755 BP; 386 A; 425 C; 437 G; 507 T; 0 other;
 Query Match 62.1%; Score 17.4; DB 23; Length 1755;
 Best Local Similarity 77.8%; Pred. NO. 1.5e+02;
 Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 OY 2 GTTCGTGGAGATGCGTCGATGCTT 28
 ||||||||| ||||| ||| ||
 Db 219 GTTCGTGGAGAACTCGTCTGGGATT 245

Search completed: July 10, 2003, 19:52:53
 Job time : 67.0831 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 12.5213 Seconds
(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631B-9

Perfect score: 28
Sequence: 1 cgttcgtgggtatgctcgtcgtgtgt 28

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2.6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2.6/ptodata/1/ina/PCIOS.COMB.seq:*
6: /cgn2.6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.6	62.9	2476	US-09-221-017B-876	Sequence 876, App
2	17.4	62.1	346	US-09-060-756-256	Sequence 256, App
3	17.4	62.1	4403765	US-09-103-840A-2	Sequence 2, Appl
4	17.4	62.1	4411529	US-09-103-840A-1	Sequence 1, Appl
5	17	60.7	1290	US-08-827-291A-1	Sequence 1, Appl
6	17	60.7	2696	US-07-961-522-1	Sequence 1, Appl
7	17	60.7	2696	US-08-217-438-1	Sequence 1, Appl
8	17	60.7	2696	US-08-321-978-1	Sequence 1, Appl
9	17	60.7	2696	US-08-710-584-1	Sequence 1, Appl
10	16.8	60.0	2418	US-09-388-743-25	Sequence 25, Appl
11	16.8	60.0	2781	US-09-342-681C-116	Sequence 116, App
12	16.8	60.0	4235	US-09-342-681C-18	Sequence 18, App
13	16.8	60.0	5476	US-09-221-017B-147	Sequence 147, App
14	16.6	59.3	367	US-08-936-165A-217	Sequence 217, App
15	16.6	59.3	948	US-09-134-001C-724	Sequence 724, App
16	16.6	58.6	405	US-08-806-597A-7	Sequence 7, Appl
17	16.4	58.6	405	US-08-970-428A-7	Sequence 7, Appl
18	16.4	58.6	1050	US-08-806-597A-13	Sequence 13, Appl
19	16.4	58.6	1050	US-08-970-428A-13	Sequence 13, Appl
20	16.4	58.6	2588	PCR-US91-02954-1	Sequence 1, Appl
21	16.4	58.6	2608	US-07-725-083-1	Sequence 1, Appl
22	16.2	57.9	1817	US-08-538-816A-3	Sequence 3, Appl
23	16.2	57.9	1817	US-09-076-651-3	Sequence 3, Appl
24	16.2	57.9	1817	US-09-208-394-3	Sequence 3, Appl
25	16.2	57.9	1844	US-08-538-816A-10	Sequence 10, Appl
26	16.2	57.9	1844	US-09-076-651-10	Sequence 10, Appl
27	16.2	57.9	1844	US-09-208-394-10	Sequence 10, Appl

28	16.2	57.9	1894	US-08-538-816A-8	Sequence 8, Appl
29	16.2	57.9	1894	US-09-076-651-8	Sequence 8, Appl
30	16.2	57.9	1894	US-09-208-394-8	Sequence 8, Appl
31	16	57.1	1799	US-08-687-590-58	Sequence 58, Appl
32	16	57.1	1988	US-08-961-527-268	Sequence 268, App
33	16	57.1	3073	US-08-975-762-41	Sequence 41, Appl
34	16	57.1	3073	US-09-295-028-41	Sequence 41, Appl
35	16	57.1	3073	US-09-106-582-41	Sequence 41, Appl
36	16	57.1	6122	US-08-403-545-1	Sequence 1, Appl
37	16	57.1	6122	US-08-404-381-1	Sequence 1, Appl
38	16	57.1	7032	US-08-149-097D-24	Sequence 24, Appl
39	16	57.1	7032	US-08-949-386-24	Sequence 24, Appl
40	16	57.1	7032	US-08-450-562-24	Sequence 24, Appl
41	16	57.1	7032	US-08-984-709A-24	Sequence 24, Appl
42	16	57.1	7032	US-08-450-272-24	Sequence 24, Appl
43	16	57.1	7089	US-08-949-386-25	Sequence 25, Appl
44	16	57.1	7089	US-08-450-562-25	Sequence 25, Appl
45	16	57.1	7089	US-08-984-709A-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-09-221-017B-876/C
Sequence 876, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: ROSS, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESS: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Montoy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 876:
SEQUENCE CHARACTERISTICS:
LENGTH: 2476 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1...2476
US-09-221-017B-876

Query Match
Best Local Similarity 62.1%; Score 17.6; DB 4; Length 2476;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 5 CGTGGATAGTCGCGTATGCTGT 28
996 CGTGGATGTCGTCACCGAGT 973

RESULT 2

US-09-060-756-256/C
Sequence 256, Application US/09060756
Patent No. 6183957
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: Cole, Stewart
APPLICANT: Buchleser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060,756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 256
LENGTH: 346
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: unsure
LOCATION: (various positions within the sequence)
OTHER INFORMATION: Applicants are uncertain of bases designated as "n".
US-09-060-756-256

Query Match
Best Local Similarity 62.1%; Score 17.4; DB 4; Length 346;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GTTCGTGGATAGTCGCTC 20
Db 253 GTTCGTGGATAGTCGCTC 235

RESULT 3

US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765

TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match
Best Local Similarity 62.1%; Score 17.4; DB 4; Length 4403765;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CGTTCGTGGATAGTCGCGTATGCTGT 27
Db 3836409 CGTTCGTGGATAGTCGCGTATGCTGT 3836435

RESULT 4

US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match
Best Local Similarity 62.1%; Score 17.4; DB 4; Length 4411529;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CGTTCGTGGATAGTCGCGTATGCTGT 27
Db 3850511 CGTTCGTGGATAGTCGCGTATGCTGT 3850537

RESULT 5

US-08-827-291A-1
Sequence 1, Application US/08827291A
Patent No. 5874243
GENERAL INFORMATION:
APPLICANT: Macina, Roberto
APPLICANT: Sathe, Ganesh
TITLE OF INVENTION: NOVEL OLRCC15 RECEPTOR
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
APPLICATION DATA:
APPLICATION NUMBER: US/08/827,291A
FILING DATE: 28-MAR-1997

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: KING, William T
REGISTRATION NUMBER: 30,954
REFERENCE/DOCKET NUMBER: GP50001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5015
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1290 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-827-291A-1

Query Match 60.7%; Score 17; DB 2; Length 1290;
Best Local Similarity 80.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGTTCGTGGATAGTCGTCATGCT 25
|||||
DB 408 CCTTCATGGGAACCTCTCATGG 432

RESULT 6
US-07-961-522-1
Sequence 1, Application US/07961522
Patent No. 5417971
GENERAL INFORMATION:
APPLICANT: POTTER, ANDREW A.
APPLICANT: GERLACH, GERALD F.
APPLICANT: WILSON, PHILIP J.
APPLICANT: ROSSI-CAMPOS, AMALIA
TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
TITLE OF INVENTION: PLEUROSPERMIONIAE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROBERTA L. ROBINS
STREET: 635 BRYANT STREET
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/961,522
FILING DATE: 19921015
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9000-0015.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 617-8999
TELEFAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:

NAME/KEY: CDS
LOCATION: 333..1973
US-07-961-522-1

Query Match 60.7%; Score 17; DB 1; Length 2696;
Best Local Similarity 80.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 TTCGTGGATAGTCGTCATGCTGT 27
|||||
DB 2267 TTCGTGGGATAGTCGTCATGCTGT 2291

RESULT 7
US-08-217-438-1
Sequence 1, Application US/08217438
Patent No. 5521072
GENERAL INFORMATION:
APPLICANT: Potter, Andrew A.
APPLICANT: Gerlach, Gerald F.
APPLICANT: Wilson, Philip J.
APPLICANT: Rossi-Campos, Amalia
TITLE OF INVENTION: ACTINOBACILLUS PLEUROSPERMIONIAE
TITLE OF INVENTION: TRANSFERIN BINDING PROTEINS AND USES THEREOF
NUMBER OF SEQUENCES: 11
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/217,438
FILING DATE: 22-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Robins, Roberta L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-617-8999
TELEFAX: 415-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1973
US-08-217-438-1

Query Match 60.7%; Score 17; DB 1; Length 2696;
Best Local Similarity 80.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 TTCGTGGATAGTCGTCATGCTGT 27
|||||
DB 2267 TTCGTGGGATAGTCGTCATGCTGT 2291

RESULT 8
US-08-321-978-1
Sequence 1, Application US/08321978
Patent No. 5801018

GENERAL INFORMATION:
APPLICANT: POTTER, ANDREW A.
APPLICANT: GERLACH, GERALD F.
APPLICANT: WILSON, PHILIP J.
APPLICANT: ROSSI-CAMPOS, ANALLA
TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROBERTA L. ROBINS
STREET: 285 HAMILTON AVE, SUITE 200
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/321,978
FILING DATE: 12-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-3400
FAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1973
US-08-321-978-1

Query Match 60.7%; Score 17; DB 1; Length 2696;
Best Local Similarity 80.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TTCGTGGAGATGCGTCATGATGTT 27
Db 2267 TTCGTGGGTAAGATCGTAATCGTGT 2291

RESULT 9
US-08-710-584-1
Sequence 1, Application US/08710584
Patent No. 5876725
GENERAL INFORMATION:
APPLICANT: POTTER, ANDREW A.
APPLICANT: GERLACH, GERALD F.
APPLICANT: WILSON, PHILIP J.
APPLICANT: ROSSI-CAMPOS, ANALLA
TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROBERTA L. ROBINS
STREET: 285 HAMILTON AVE, SUITE 200
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/710,584
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/321,978
FILING DATE: 12-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-3400
FAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1973
US-08-710-584-1

Query Match 60.7%; Score 17; DB 2; Length 2696;
Best Local Similarity 80.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TTCGTGGAGATGCGTCATGATGTT 27
Db 2267 TTCGTGGGTAAGATCGTAATCGTGT 2291

RESULT 10
US-09-388-743-25
Sequence 25, Application US/09388743
Patent No. 6423886
GENERAL INFORMATION:
APPLICANT: Singletary, George
APPLICANT: Zhou, Ian
TITLE OF INVENTION: Use in the Production of New Starches
FILE REFERENCE: 1144
CURRENT APPLICATION NUMBER: US/09/388,743
CURRENT FILING DATE: 1999-09-02
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 25
LENGTH: 2418
TYPE: DNA
ORGANISM: Typha latifolia
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(2418)
US-09-388-743-25

Query Match 60.0%; Score 16.8; DB 4; Length 2418;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CTTGCTGGAGATGCGTCATGATGTT 28
Db 1656 CATGCTTGATGTGACGATCATGATGTT 1683

RESULT 11
US-09-342-681C-116/C
Sequence 116, Application US/09342681C

Patent No. 6355782
GENERAL INFORMATION:
APPLICANT: Zonana et al.
TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
FILE REFERENCE: 52978
CURRENT APPLICATION NUMBER: US/09/342,681C
CURRENT FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 60/092,279
PRIOR FILING DATE: 1998-07-09
PRIOR APPLICATION NUMBER: 60/112,366
PRIOR FILING DATE: 1998-12-15
NUMBER OF SEQ ID NOS: 123
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 116
LENGTH: 2781
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (528)..(2756)
OTHER INFORMATION: n is a, c, t or g
US-09-342-681C-116

Query Match 60.0%; Score 16.8; DB 4; Length 2781;
Best Local Similarity 75.0%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 CGTTCGTGGAGTACCGTCATGCGTGT 28
DB 2010 CGTTCGTGGAGTACCGTCATGCGTGT 1983

RESULT 12
US-09-342-681C-18/C
Sequence 18, Application US/09342681C
Patent No. 6355782
GENERAL INFORMATION:
APPLICANT: Zonana et al.
TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
FILE REFERENCE: 52978
CURRENT APPLICATION NUMBER: US/09/342,681C
CURRENT FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 60/092,279
PRIOR FILING DATE: 1998-07-09
PRIOR APPLICATION NUMBER: 60/112,366
PRIOR FILING DATE: 1998-12-15
NUMBER OF SEQ ID NOS: 123
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 18
LENGTH: 4235
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (433)..(1779)
US-09-342-681C-18

Query Match 60.0%; Score 16.8; DB 4; Length 4235;
Best Local Similarity 75.0%; Pred. No. 48;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 CGTTCGTGGAGTACCGTCATGCGTGT 28
DB 3466 CGTTCGTGGAGTACCGTCATGCGTGT 3439

RESULT 13
US-09-221-017B-147
Sequence 147, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM type: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 147:
SEQUENCE CHARACTERISTICS:
LENGTH: 3476 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: circular
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...5476
US-09-221-017B-147

Query Match 60.0%; Score 16.8; DB 4; Length 5476;
Best Local Similarity 75.0%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 CGTTCGTGGAGTACCGTCATGCGTGT 28
DB 5013 CATTCGTGGAGAGAGCCGCGATGCGAT 5040

RESULT 14
US-08-936-165A-217
Sequence 217, Application US/08936165A
Patent No. 6348582
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Lonetto, Michael
APPLICANT: Nicholas, Richard

APPLICANT: Pratt, Julie
 APPLICANT: Reichard, Richard
 APPLICANT: Rosenberg, Martin
 APPLICANT: Ward, Judith
 TITLE OF INVENTION: No. 6348582a1 Prokaryotic Polynucleotides,
 TITLE OF INVENTION: Polypeptides and Their Uses
 NUMBER OF SEQUENCES: 534
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Smithline Beecham Corporation
 STREET: 709 Swedeland Road
 CITY: King of Prussia
 STATE: PA
 COUNTRY: USA
 ZIP: 19406-0939
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/936,165A
 FILING DATE: 24-SEP-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/027,032
 FILING DATE: 24-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Gimm1, Edward R.
 REGISTRATION NUMBER: 38,891
 REFERENCE/DOCKET NUMBER: P50549
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 610-270-4478
 TELEFAX: 610-270-5090
 TELEX:
 INFORMATION FOR SEQ ID NO: 217:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 367 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: genomic DNA
 US-08-936-165A-217
 Query Match 59.3% Score 16.6; DB 4; Length 367;
 Best Local Similarity 82.6% Pred. No. 40;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 4 TCGTGGATAGTCCGTCATGGTG 26
 DB 251 TCTTGGATACCTCCGTTAGGCTG 273
 RESULT 15
 US-09-134-001C-724/c
 Sequence 724, Application US/09134001C
 Patent No. 6380370
 GENERAL INFORMATION:
 APPLICANT: Lynn Doucelle-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134,001C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 724
 LENGTH: 948
 TYPE: DNA
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-724

Query Match 59.3% Score 16.6; DB 4; Length 948;
 Best Local Similarity 82.6% Pred. No. 47;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 CGTTCGTGGATAGTCCGTCATG 23
 DB 106 CGTTCGTGGATAGTCCGTCATG 84
 Search completed: July 10, 2003, 20:28:02
 Job time : 33.5213 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:53:16 ; Search time 76.6382 Seconds
(without alignments)
576.891 Million cell updates/sec

Title: US-10-087-631B-9

Perfect score: 28

Sequence: 1 cgttcgtgagatagtcgcatgtgtt 28

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	28	100.0	28	US-10-087-631B-9	Sequence 9, Appl1
2	28	100.0	241	US-10-087-631B-12	Sequence 12, Appl1
3	28	100.0	943	US-10-087-631B-11	Sequence 11, Appl1
4	20	71.4	1533	US-10-156-761-6429	Sequence 6429, Ap
5	20	71.4	9025608	US-10-156-761-1	Sequence 1, Appl1
6	17.4	62.1	1572	US-10-156-761-5471	Sequence 5471, Ap
7	17.4	62.1	8549	US-08-781-986A-5	Sequence 5, Appl1
8	17.4	62.1	684973	US-09-263-959-1	Sequence 1, Appl1
9	17.4	62.1	9025608	US-10-156-761-1	Sequence 1, Appl1
10	17.2	61.4	1247	US-09-880-107-3874	Sequence 3874, Ap
11	17.2	61.4	1320	US-10-098-841-33	Sequence 33, Appl1
12	17	60.7	493	US-09-796-692-4625	Sequence 4625, Ap
13	17	60.7	493	US-10-040-862-4625	Sequence 4625, Ap
14	17	60.7	924	US-10-156-761-1019	Sequence 1019, Ap
15	17	60.7	1290	US-09-826-508-23	Sequence 23, Appl1
16	16.8	60.0	1338	US-09-815-242-7956	Sequence 7956, Ap
17	16.8	60.0	2781	US-09-729-658B-116	Sequence 116, Ap
18	16.8	60.0	3623	US-09-918-909-23	Sequence 23, Appl1
19	16.8	60.0	4235	US-09-729-658B-18	Sequence 18, Appl1

20	16.6	59.3	367	10	US-09-939-980-217	Sequence 217, App
21	16.6	59.3	567	10	US-09-864-761-16516	Sequence 16516, A
22	16.4	58.6	428	10	US-09-954-456-1334	Sequence 1334, Ap
23	16.4	58.6	583	9	US-10-106-698-3533	Sequence 3533, Ap
24	16.4	58.6	633	12	US-10-029-079-4	Sequence 4, Appl1
25	16.4	58.6	896	9	US-10-007-280A-125	Sequence 125, App
26	16.4	58.6	998	9	US-10-007-280A-126	Sequence 126, App
27	16.4	58.6	1017	9	US-10-198-846-11216	Sequence 11216, A
28	16.4	58.6	1038	10	US-09-974-300-439	Sequence 439, App
29	16.4	58.6	1131	10	US-09-925-302-26	Sequence 26, Appl1
30	16.4	58.6	1325	9	US-09-925-299-185	Sequence 185, App
31	16.4	58.6	1325	10	US-09-925-299-185	Sequence 185, App
32	16.4	58.6	1328	9	US-10-082-830-23	Sequence 23, Appl1
33	16.4	58.6	1467	9	US-10-156-761-1138	Sequence 1138, Ap
34	16.4	58.6	1638	9	US-10-270-333-170	Sequence 170, App
35	16.4	58.6	1638	10	US-09-969-147-212	Sequence 212, App
36	16.4	58.6	1830	10	US-09-925-300-224	Sequence 224, App
37	16.4	58.6	1979	10	US-09-925-300-224	Sequence 224, App
38	16.4	58.6	2182	9	US-10-091-572-816	Sequence 816, App
39	16.4	58.6	2182	9	US-09-764-891-9240	Sequence 9240, Ap
40	16.4	58.6	2244	7	US-08-781-986A-270	Sequence 270, App
41	16.4	58.6	2723	9	US-10-177-293-372	Sequence 372, App
42	16.4	58.6	2921	10	US-09-925-301-223	Sequence 223, App
43	16.4	58.6	3395	9	US-10-198-846-13880	Sequence 13880, A
44	16.4	58.6	3809	9	US-10-273-680-1	Sequence 1, Appl1
45	16.4	58.6	3957	9	US-10-270-333-169	Sequence 169, App

ALIGNMENTS

RESULT 1
US-10-087-631B-9
Sequence 9, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
TITLE OF INVENTION: CONTROL
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 9
LENGTH: 28
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: 5778pc primer parallel-
US-10-087-631B-9

Query Match 100.0%; Score 28; DB 9; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OR
1 CGTTCGTGAGTACTCGTCATGTGTT 28
DB
1 CGTTCGTGAGTACTCGTCATGTGTT 28
RESULT 2
US-10-087-631B-12/c
Sequence 12, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
TITLE OF INVENTION: CONTROL
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 12
LENGTH: 241
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Amplicon derived from QS (pc
US-10-087-631B-12
OTHER INFORMATION: using the primers ST280pc and ST778pc

Query Match
Best Local Similarity 100.0%; Score 28; DB 9; Length 241;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTTCGTGGATAGTCCGTCATGCTGT 28
|||||
DB 241 CGTTCGTGGATAGTCCGTCATGCTGT 214

RESULT 3
US-10-087-631B-11/c
Sequence 11, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11
LENGTH: 943
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: QS (pc) HCV being parallel-
US-10-087-631B-11
OTHER INFORMATION: complement to according region of HCV type 1 genome

Query Match
Best Local Similarity 100.0%; Score 28; DB 9; Length 943;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTTCGTGGATAGTCCGTCATGCTGT 28
|||||
DB 297 CGTTCGTGGATAGTCCGTCATGCTGT 270

RESULT 4
US-10-156-761-6429/c
Sequence 6429, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 6429

LENGTH: 1533
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1533)
US-10-156-761-6429

Query Match
Best Local Similarity 71.4%; Score 20; DB 9; Length 1533;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGTTCGTGGATAGTCCGTCATGCTGT 28
|||||
DB 196 CGATCGTGTGACGCTCGCTTGTGCTT 169

RESULT 5
US-10-156-761-1
Sequence 1, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match
Best Local Similarity 71.4%; Score 20; DB 9; Length 9025608;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGTTCGTGGATAGTCCGTCATGCTGT 28
|||||
DB 7731103 CGATCGTGTGACGCTCGCTTGTGCTT 7731130

RESULT 6
US-10-156-761-5471
Sequence 5471, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 5471
LENGTH: 1572
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1572)
US-10-156-761-5471

Query Match 62.1%; Score 17.4; DB 9; Length 1572;
Best Local Similarity 77.8%; Pred. No. 99;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GTTCGTGGGATAGTCGTCATGCGTGT 28
DB 1143 GTTCGTGGGATAGTCGTCATGCGTGT 1169

RESULT 7
US-08-781-986A-5/C
Sequence 5, Application US/08781986A
Publication No. US20030054436A1

GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8549 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-08-781-986A-5

Query Match 62.1%; Score 17.4; DB 7; Length 8549;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GTTCGTGGGATAGTCGTCATGCGTGT 28
DB 6487 GTTCATAGGCTAGTCTGTAATAGTGT 6461

RESULT 8
US-09-263-959-1

Sequence 1, Application US/09263959
Patent No. US20020150891A1

GENERAL INFORMATION:
APPLICANT: Hood, Leroy E.
APPLICANT: Rowen, Lee
APPLICANT: Koop, Ben F.
TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH U
NUMBER OF SEQUENCES: 1279
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263,959
FILING DATE: 05-MAR-1999
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: McMaisters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 920010.426C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 684973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-09-263-959-1

Query Match 62.1%; Score 17.4; DB 10; Length 684973;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GTTCGTGGGATAGTCGTCATGCGTGT 28
DB 513413 GTTCGTGGGATAGTCGTCATGCGTGT 513439

RESULT 9
US-10-156-761-1/C

Sequence 1, Application US/10156761
Publication No. US20030119018A1

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIRAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1

LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc.feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match
Best Local Similarity 62.1%; Score 17.4; DB 9; Length 9025608;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 GTTCGTGGATAGTCCTCATGCTGT 28
DB 6658937 GTTCGTGGATAGTCCTCATGCTGT 6658911

RESULT 10
US-09-880-107-3874/C
Sequence 3874, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
PRIOR FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3874
LENGTH: 1247
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 Z29481
US-09-880-107-3874

Query Match
Best Local Similarity 61.4%; Score 17.2; DB 10; Length 1247;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 TGGGATAGTCCTCATGCTGT 28
DB 182 TGTGATAGTCCTCATGCTGT 161

RESULT 11
US-10-098-841-33/C
Sequence 33, Application US/10098841
Publication No. US20020197679A1
GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Xu, Chongjun
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungqing
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhao, Qing A.
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: Wehrman, Tom
APPLICANT: Zhang, Jie

APPLICANT: Qian, Xiaohong B.
APPLICANT: Dmanac, Radoje T.
TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
FILE REFERENCE: 784CIP2
CURRENT APPLICATION NUMBER: US/10/098,841
PRIOR FILING DATE: 2002-03-13
PRIOR APPLICATION NUMBER: 09/598,042
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 331
SOFTWARE: pl_FL_genes Version 1.0
SEQ ID NO 33
LENGTH: 1320
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (103)..(963)
US-10-098-841-33

Query Match
Best Local Similarity 61.4%; Score 17.2; DB 9; Length 1320;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 TGGGATAGTCCTCATGCTGT 28
DB 244 TGTGATAGTCCTCATGCTGT 223

RESULT 12
US-09-796-692-4625/C
Sequence 4625, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
FILE REFERENCE: 2077,001200
CURRENT APPLICATION NUMBER: US/09/796,692
PRIOR FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/223,378
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4625

```

; LENGTH: 493
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-4625

```

Query Match	60.7%	Score 17	DB 9	Length 493
Best Local Similarity	80.0%	Pred. No	14e+02	
Matches	20	Conservative	0	Mismatches 5
				Indels 0
				Gaps 0

```

QY      4 TCGTGGATAGTCCGTCATGGTGT 28
          ||||| ||| | ||||| ||| |
Db      48 TCGTGAATGCTATGTCATGGTGT 24

```

RESULT 13
US-10-040-862-4625/c
: Sequence 4625, Application US/10040862
: Publication NO. US20030078396A1
: GENERAL INFORMATION:
: APPLICANT: Gaiger, Alexander
: APPLICANT: Aligate, Paul A.
: APPLICANT: Mannion, Jane
: APPLICANT: Retter, Marc
: APPLICANT: Corixa Corporation
: TITLE OF INVENTION: Compositions and Methods for the Detection,
: TITLE OF INVENTION: Hematological Malignancies
: TITLE OF INVENTION: Hematological Malignancies

ORGANISM: Homo sapiens
US-10-040-862-4625

Query Match	60.7%	Score 17	DB 9	Length 493
Best Local Similarity	80.0%	Pred. No. 1.4e+02		
Matches 20	Conservative 0	Mismatches 5	Indels 0	Gaps 0

RESULT 14

US-10-156-761-1019/c
; Sequence 1019, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

Query Match	60.7%	Score 17	DB 9	Length 924
Best Local Similarity	80.0%	Pred. No.	1.5e+02	
Matches 20	Conservative 0	Mismatches 5	Indels 0	Gaps 0

```

oy      2 GTTCGTGGGATAGTCCGTCATGGTG 26
          ||||| | | ||||| | | | |
Db      795 GTTCGGGGGTAGTCCGTCGTGATG 771

```

```

RESULT 15
US-09-826-508-23
; Sequence 23, Application US/09826508
; Patent No. US20010025099A1
; GENERAL INFORMATION:
; APPLICANT: Nabil Elshourbagy
; APPLICANT: Lisa Vawter
; TITLE OF INVENTION: G Protein-Coupled Receptor Polypeptides
; TITLE OF INVENTION: and Polynucleotides
; FILE REFERENCE: Gp-70744USB
; CURRENT APPLICATION NUMBER: US/09/826,508
; CURRENT FILING DATE: 2001-04-05
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-826-508-23

```

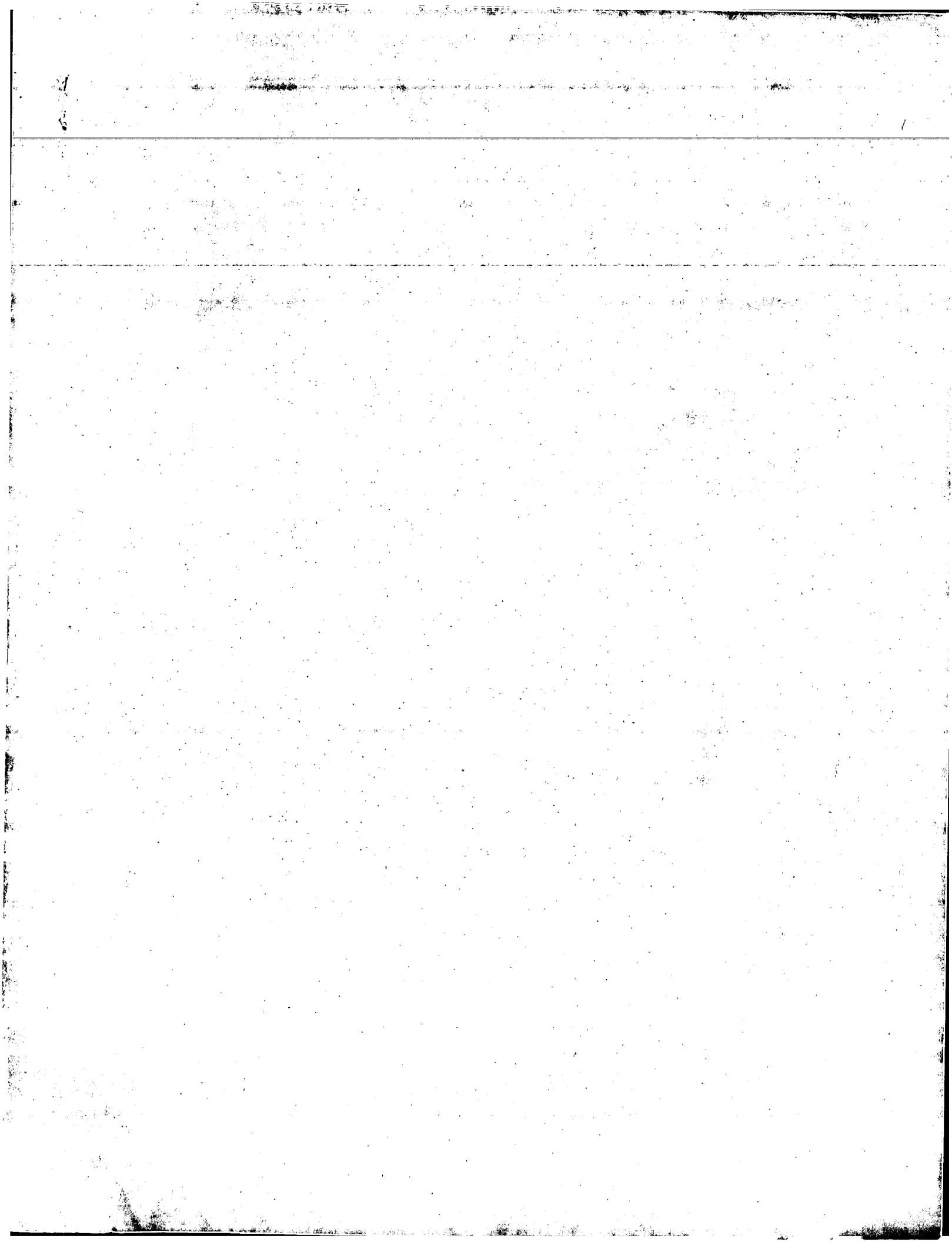
Query Match..	60.7%;	Score 17;	DB 10;	Length 1290;
Best Local Similarity	80.0%;	Pred. No. 1.5e+02;		
Matches 20;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;

```

OY      1 CGTTCGTGGGATAGTCCGCATGTT 25
          | ||| |||| | | ||||| ||
Db      408 CCTTCATGGGAACCTCTGTCATGGT 432

```

Search completed: July 11, 2003, 15:02:47
Job time : 117.638 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:13:45 ; Search time 503.434 Seconds
(without alignments)
900.761 Million cell updates/sec

Title: US-10-087-631b-9

Perfect score: 28
Sequence: 1 cgttcgtgagatagtcgcatggtt 28

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	20	71.4	475	13	BI611725
c 2	19	67.9	406	10	AV732708
c 3	19	67.9	590	17	A2567844
c 4	19	67.9	649	13	BMS97959
c 5	19	67.9	667	9	AI055301
c 6	19	67.9	900	12	BF620238

Result No.	Score	Query Match	Length	DB ID	Description
c 7	18.6	66.4	521	17	AQ350115
c 8	18.6	66.4	549	17	BH282580
c 9	18.6	66.4	552	17	AQ776894
c 10	18.6	66.4	560	17	A2397654
c 11	18.6	66.4	635	17	BH265153
c 12	18.6	66.4	640	17	A2291296
c 13	18.6	66.4	736	12	BG293648
c 14	18.6	66.4	763	17	BH267142
c 15	18.6	66.4	805	17	BH279359
c 16	18.6	66.4	865	17	AQ740087
c 17	18.6	66.4	868	13	BI761876
c 18	18.4	65.7	131	10	BE148040
c 19	18.4	65.7	247	12	BG410222
c 20	18.4	65.7	265	10	BG410420
c 21	18.4	65.7	289	10	BH348944
c 22	18.4	65.7	325	10	AM566217
c 23	18.4	65.7	400	10	AM528213
c 24	18.4	65.7	480	9	AA820730
c 25	18.4	65.7	482	14	BQ996871
c 26	18.4	65.7	513	12	BE978187
c 27	18.4	65.7	529	17	AQ913537
c 28	18.4	65.7	533	13	BI167853
c 29	18.4	65.7	541	13	BI163484
c 30	18.4	65.7	560	13	BI170216
c 31	18.4	65.7	573	13	BI229543
c 32	18.4	65.7	593	13	BI632278
c 33	18.4	65.7	605	9	AT520529
c 34	18.4	65.7	613	9	AA568072
c 35	18.4	65.7	620	14	BQ988356
c 36	18.4	65.7	652	14	BQ999054
c 37	18.4	65.7	662	13	BI636271
c 38	18.4	65.7	752	17	A2132792
c 39	18.4	65.7	769	17	A2132929
c 40	18.4	65.7	812	12	BG435911
c 41	18.4	65.7	1064	14	BQ048851
c 42	18.2	65.0	757	17	AG185522
c 43	18.2	65.0	846	12	BF138165
c 44	18.2	65.0	1033	12	BG390990
c 45	18.2	64.3	166	9	AV058751

ALIGNMENTS

RESULT 1
LOCUS BI611725/c 475 bp mRNA linear EST 07-SEP-2001
DEFINITION RH18118.5prlme RH Drosophila melanogaster normalised Head pfc1-1
Drosophila melanogaster cDNA clone RH18118.5 similar to sqd:
Fban0017791 GO:RNA binding (GO:0003723); RNA binding (GO:0003723)]
located on: 3R 87F-87F7; 08/17/2001, mRNA sequence.

ACCESSION BI611725.1 GI:15507250
VERSION BI611725.1

KEYWORDS fruit fly
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 475)
J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Mista, S., Mungall, C., Nunoo, J., Paclib, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celisner, S. and Rubin, G.M.

TITLE BDGP/HMI RH Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
hit genomic AEO03701: arm:3R (935746,956163)
estimated-cyt:87F4-87F4: 08/17/2001
Plate: RH 181 row: B Column: 6
High quality sequence stop: 291.

FEATURES

Location/Qualifiers
1..475

ORGANISM="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RH1818"
/clone_lib="RH Drosophila melanogaster normalized Head
PFLC-1"
/sex="male and female"
/dev_stage="Adult"
/lab_host="DHS-alpha Tona"
/note="Organ: head; Vector: pFLC1; Site:1; XhoI; Site:2;
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid CDNA library."
BASE COUNT
150 a 108 c 110 g 107 t

Query Match

Best Local Similarity 71.4%; Score 20; DB 13; Length 475;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db

1 CGTTCGTGGATAGTCGTCATGCGTGT 28
347 CCTCTGGGATAGACCTTCATGCGTGT 320

RESULT 2
AV732708 406 bp mRNA linear EST 17-OCT-2000
LOCUS
DEFINITION
AV732708 HTF Homo sapiens CDNA clone HFBCH01 5', mRNA sequence.
AV732708
AV732708.1 GI:10850253
EST.
SOURCE
KEYWORDS
ORGANISM
human.

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 406)
Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,
Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu
Chen, J., Chen, Z., and Han, Z.
Homo sapiens CDNA HTF clones
Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzeg@hgc.sh.cn

TITLE

JOURNAL
COMMENT

FEATURES

Location/Qualifiers
1..406

ORGANISM="Homo sapiens"
/db_xref="taxon:9606"
/clone="HFBCH01"
/clone_lib="HTF"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site:1: EcoRI; Site:2:
XhoI"
BASE COUNT
153 a 50 c 65 g 138 t

Query Match

Best Local Similarity 67.9%; Score 19; DB 10; Length 406;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Matches

22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db

1 CGTTCGTGGATAGTCGTCATGCGTGT 27
376 CGATGCGGTATAGTACGTCATGCTT 402

RESULT 3
A2567844 590 bp DNA linear GSS 15-MAY-2001
LOCUS
DEFINITION
A2567844 239PvH07 Pv MBN #30 Plasmodium vivax genomic 3', DNA sequence.
VERSION
A2567844.1 GI:13976352
KEYWORDS
SOURCE
ORGANISM
malaria parasite P. vivax.
Plasmodium vivax
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 590)
Carlton, J.M., R. and Dame, J.B.
The Plasmodium vivax and P. berghei gene sequence tag projects
Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.ufl.edu
Seq primer: M13(-20) forward
Class: Shotgun.

REFERENCE

1 (bases 1 to 590)
/organism="Plasmodium vivax"
/strain="Salvador I (Collins, W. 1972. J. Parasitol. 69,
497-598)."
/db_xref="taxon:5855"
/clone_lib="Pv MBN #30"
/dev_stage="asexual blood forms"
/lab_host="Salintri boliviensis"
/note="Vector: pBluescript SK(+)
excised from lambda ZAP. Site:1: EcoR V; Site:2: EcoR V;
Host leukocytes were extracted from P. vivax infected
blood using the following methods: first, infected blood
per 10 ml blood. Then blood was passed over a column of
acid washed 0.1 mm glass beads, then through a Plasmidipur
Whatman CF11 powder (1:2 ratio volume of blood to CF11),
cushion. Purified DNA was digested with mung bean nuclease
(Vernick, K.D., Imberiski, R.B., and McCutchan, T.F. 1988.
Nucleic Acids Research 16:6883-6896). Digested DNA was
blunt-ended using T4 DNA Polymerase and size fractionated
over a Sepharose CL-2B column. Fractions in the size range
300bp-4kb were ligated into the Eco RV site of pBluescript
SK(+), and E. coli XL-10 Gold transformed with the
ligation mixture."

FEATURES

Location/Qualifiers
1..590

ORGANISM="Plasmodium vivax"
/db_xref="taxon:5855"
/clone_lib="Pv MBN #30"
/dev_stage="asexual blood forms"
/lab_host="Salintri boliviensis"
/note="Vector: pBluescript SK(+)
excised from lambda ZAP. Site:1: EcoR V; Site:2: EcoR V;
Host leukocytes were extracted from P. vivax infected
blood using the following methods: first, infected blood
per 10 ml blood. Then blood was passed over a column of
acid washed 0.1 mm glass beads, then through a Plasmidipur
Whatman CF11 powder (1:2 ratio volume of blood to CF11),
cushion. Purified DNA was digested with mung bean nuclease
(Vernick, K.D., Imberiski, R.B., and McCutchan, T.F. 1988.
Nucleic Acids Research 16:6883-6896). Digested DNA was
blunt-ended using T4 DNA Polymerase and size fractionated
over a Sepharose CL-2B column. Fractions in the size range
300bp-4kb were ligated into the Eco RV site of pBluescript
SK(+), and E. coli XL-10 Gold transformed with the
ligation mixture."

BASE COUNT

142 a 138 c 155 g 155 t

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db

1 CGTTCGTGGATAGTCGTCATGCGTGT 27
504 CGTTCGTGGAGCGCCGTGATGCTT 530

LOCUS

BMS97959 649 bp mRNA linear EST 25-FEB-2002

```

Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: SP030 (AACAGCATGACCATGATTA)
High quality sequence stop: 325.
Location/Qualifiers
1. 667
/organism="Gossypium hirsutum"
/db_xref="taxon:3635"
/cclone="coau0003018"
/cclone_id="Cotton Boll1 Abscission Zone cDNA Library"
/lab_host="XLI-Blue MRF"
/note="Vector: Uni-ZAP XR; Site 1: EcoRI; Site 2: XhoI;
This is a Uni-ZAP XR custom cDNA library made by
Stratagene (U.S.A.: 1-800-424-5444); Stratagene cat.
#837201."
BASE COUNT      176 a      141 c      149 g      199 t      2 others
ORIGIN
Query Match      67.9%; Score 19; DB 9; Length 667;
Best Local Similarity 81.5%; Pred. No. 4.7e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY      1 CGTTCGTGGGATAGTCGTCATGCTGT 27
Db      45 CGCTGTGGGATGTCTTCTCTGTGT 71
RESULT 6
BF620238/c
LOCUS      BF620238      900 bp      mRNA      linear      EST 22-OCT-2001
DEFINITION      HVSMEC0019C13f Hordeum vulgare seedling shoot EST library
HVCNMA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone
HVSMEC0019C13f, mRNA sequence.
ACCESSION      BF620238
VERSION
KEYWORDS
ORGANISM
SOURCE
JOURNAL
COMMENT
TITLE
REFERENCE
AUTHORS
JOURNAL
COMMENT
FEATURES
SOURCE
/organism="Hordeum vulgare"
/db_xref="Morex"
/cclone="HVSMEC0019C13f"
/cclone_id="Hordeum vulgare seedling shoot EST library
HVCNMA0003 (Etiolated and unstressed)"
/tissue_type="Seedling shoot"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
Seeds were surface sterilized then germinated under aseptic
conditions in the dark at room temperature on filter paper

```

with water, mystatin and cefotaxime in covered crystallization dishes. Five-day old seedling shoots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million plv were in vivo excised to give Bluescript SK(-) cDNA phagemids. These steps were performed in the Tu Riverside Laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Ramo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinholz A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html>)

BASE COUNT

160 a 237 c 284 g 219 t

ORIGIN

Query Match
Best Local Similarity 67.9% Score 19; DB 12; Length 900;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1 CGTTCGTGGATGCTCATGCTGT 27
464 CGTTCGTGGATGCTCATGCTGT 438

RESULT 7
AC350115/c 521 bp DNA linear GSS 07-MAY-1999
DEFINITION RPI11-110C18.TV RPI11-11 Homo sapiens genomic clone RPI11-11-110C18,
AC350115
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
J.C.
Use of BAC End Sequences from Library RPI11-11 for Sequence-Ready
Map Building
Unpublished (1997)
Other GSSs: RPI11-110C18.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetlgr.org

JOURNAL
COMMENT
Class: BAC ends.

FEATURES

Location/Qualifiers
1..521
/organism="Homo sapiens"
/db_xref="GDB:7541921"
/db_xref="taxon:9606"

/clone="RPI11-11-110C18"
/clone_11b="RPI11-11"
/sex="Male"
/cell_type="lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPI11 Human Male BAC Library"

BASE COUNT

206 a 103 c 97 g 115 t

ORIGIN

Query Match
Best Local Similarity 66.4% Score 18.6; DB 17; Length 521;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 3 TTCTGGGATGCTCATGCTGT 27
254 TTCTGGGATGCTCATGCTGT 230

RESULT 8
BH282580/c 549 bp DNA linear GSS 30-NOV-2001
DEFINITION CH230-12818.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shartsbeyn
A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment
Unpublished (1999)
Other GSSs: CH230-12818.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaoetlgr.org

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Class: BAC ends.

Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@mai.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering_information.htm). BAC end
plate: 128 row: 1 column: 8
Seq primer: SP6
Class: BAC ends.

FEATURES
Source
Location/Qualifiers
1..549
/organism="Rattus norvegicus"
/strain="BN/SSNsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-12818"
/clone_11b="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
CHORI-230 Rat (BN/SSNsd/MCW) BAC library produced by
Pieter de Jong"

BASE COUNT

193 a 111 c 64 g 181 t

ORIGIN

Query Match
Best Local Similarity 66.4% Score 18.6; DB 17; Length 549;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 GTTCGTGGATAGTCCGTCATGCTG 26
 DB 101 GTTAGAGGAGATAGTCAGTATGCTG 77

RESULT 9
 LOCUS A0776894/c 552 bp DNA linear GSS 02-AUG-1999
 DEFINITION HS_5310.A2.B07.T7A.RPCI-11 Human Male BAC Library Homo sapiens
 genomic clone Plate=886 Col=14 Row=C, DNA sequence.

ACCESSION A0776894
 VERSION A0776894.1 GI:5679854
 KEYWORDS GSS.

ORGANISM Homo sapiens
 SOURCE human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 552)
 Mahairas,G.G., Wallace,J.C., Smith,R., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@edjlong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
 or from Research Genetics (info@resgen.com). BAC end Web Server:
 http://www.htsc.washington.edu
 Plate: 886 Row: C Column: 14
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 552.

FEATURES
 Source Location/Qualifiers

1..552
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_plate="886 Col=14 Row=C"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI methylase. Size selected DNA was cloned into the
 pBAC3.6 vector at EcoRI sites"

BASE COUNT 201 a 114 c 92 g 139 t 6 others

ORIGIN

Query Match 66.4%; Score 18.6; DB 17; Length 552;
 Best Local Similarity 84.0%; Pred. No. 6.6e-02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 TTTCGTGGATAGTCCGTCATGCTG 27
 DB 286 TTTCGTGGATAGTCCGTCATGCTG 262

RESULT 10
 LOCUS A2397654 560 bp DNA linear GSS 03-OCT-2000
 DEFINITION IM0162D13R Mouse 10kb plasmid U00C1M library Mus musculus genomic
 clone U00C1M0162D13 R, DNA sequence.
 ACCESSION A2397654
 VERSION A2397654.1 GI:10512726

KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 560)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
 84112, USA

Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00
 Plate: 0162 Row: D Column: 13
 Seq primer: CACACAGGAAACACCTAGACC
 Class: plasmid ends

High quality sequence stop: 560.

FEATURES
 Source Location/Qualifiers

1..560
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U00C1M0162D13"
 /clone_lib="Mouse 10kb plasmid U00C1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42mv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repated with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (914732114[9b]AF12902.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 118 a 106 c 156 g 180 t

ORIGIN

Query Match 66.4%; Score 18.6; DB 17; Length 560;
 Best Local Similarity 84.0%; Pred. No. 6.6e-02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 TTTCGTGGATAGTCCGTCATGCTG 27
 DB 2 TTTCGTGGATAGTCCGTCATGCTG 26

RESULT 11
 LOCUS BH265153 635 bp DNA linear GSS 30-NOV-2001
 DEFINITION CH230-55121.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
 CH230-55121, DNA sequence.
 ACCESSION BH265153
 VERSION BH265153.1 GI:17174769


```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="4502619"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT      181 a      160 c      210 g      185 t
ORIGIN

Query Match      66.4%; Score 18.6; DB 12; Length 736;
Best Local Similarity 84.0%; Pred. No. 7.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      3 TTCTGGGATGATCGCTCATGCTGT 27
        ||||| ||||| ||||| |||||
        2 TTCTGGGATGATCGCTCATGCTGT 26

RESULT 14
BH267142/c      763 bp      DNA      linear      GSS 30-NOV-2001
LOCUS      CH230-55010.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
DEFINITION      BH267142.1 GI:11719062
ACCESSION      BH267142
VERSION      BH267142.1
KEYWORDS      GSS.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
              1 (bases 1 to 763).
              Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn
              A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de
              Jong, P. and Fraser, C.M.
              Rat BAC End Sequences from Library CHORI-230 EcoRI segment
              Unpublished (1999)
              Other GSSs: CH230-55010.TV
              Contact: Shaying Zhao
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: szhao@tigr.org
              Clones are derived from the rat BAC library CHORI-230
              (http://www.chori.org/bacpac/rat230.htm). For BAC library
              availability, please contact Pieter de Jong (pdejong@tigr.org).
              Clones may be purchased from BACPAC Resources
              (http://www.chori.org/bacpac/or_ering_information.htm). BAC end
              page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
              Plate: 55 row: 0 column: 10
              Seq primer: SP6
              Class: BAC ends.

FEATURES
  source      Location/Qualifiers
              1..763
                /organism="Rattus norvegicus"
                /strain="BN/SSNhd/MCW"
                /db_xref="taxon:10116"
                /clone="CH230-55010"
                /clone_lib="CHORI-230 Segment 1"
                /sex="Female"
                /cell_type="Brain"
                /note="Vector: pTARAC2.1; Site: 1: EcoRI; Site: 2: EcoRI;
                CHORI-230 Rat (BN/SSNhd/MCW) BAC library produced by
                Pieter de Jong"

BASE COUNT      275 a      139 c      98 g      251 t
ORIGIN

```

```

Query Match      66.4%; Score 18.6; DB 17; Length 805;
Best Local Similarity 84.0%; Pred. No. 7.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      2 GTTCTGGGATGATCGCTCATGCTG 26
        ||| | ||||| || |||||
        Db      104 GTTAGAGGATGATGATGATG 80

RESULT 15
BH279359/c      805 bp      DNA      linear      GSS 30-NOV-2001
LOCUS      CH230-128C8.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
DEFINITION      BH279359
ACCESSION      BH279359
VERSION      BH279359.1 GI:117191761
KEYWORDS      GSS.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
              1 (bases 1 to 805)
              Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn
              A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de
              Jong, P. and Fraser, C.M.
              Rat BAC End Sequences from Library CHORI-230 EcoRI segment
              Unpublished (1999)
              Contact: Shaying Zhao
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: szhao@tigr.org
              Clones are derived from the rat BAC library CHORI-230
              (http://www.chori.org/bacpac/rat230.htm). For BAC library
              availability, please contact Pieter de Jong (pdejong@tigr.org).
              Clones may be purchased from BACPAC Resources
              (http://www.chori.org/bacpac/or_ering_information.htm). BAC end
              page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
              Plate: 128 row: C column: 8
              Seq primer: SP6
              Class: BAC ends.

FEATURES
  source      Location/Qualifiers
              1..805
                /organism="Rattus norvegicus"
                /strain="BN/SSNhd/MCW"
                /db_xref="taxon:10116"
                /clone="CH230-128C8"
                /clone_lib="CHORI-230 Segment 1"
                /sex="Female"
                /cell_type="Brain"
                /note="Vector: pTARAC2.1; Site: 1: EcoRI; Site: 2: EcoRI;
                CHORI-230 Rat (BN/SSNhd/MCW) BAC library produced by
                Pieter de Jong"

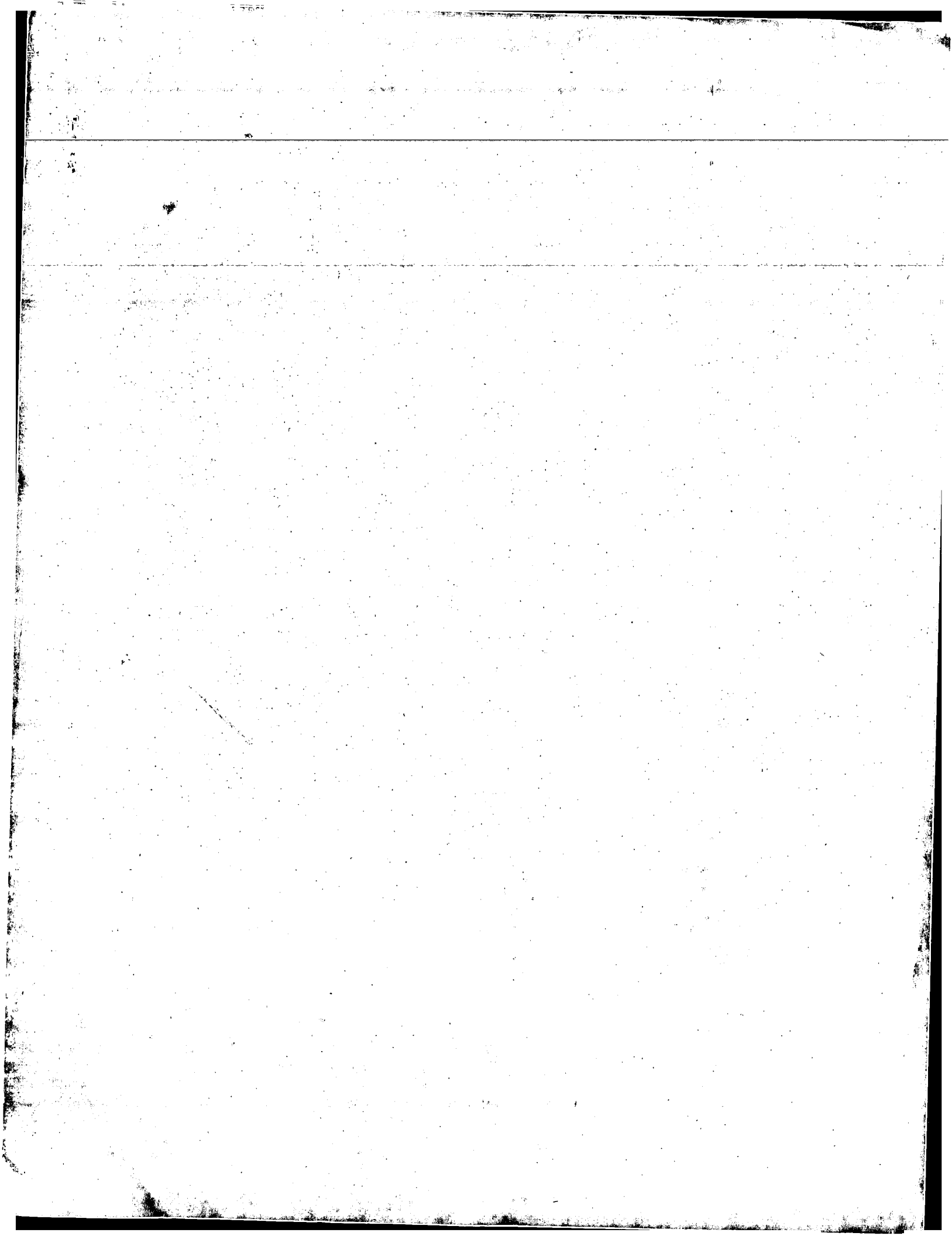
BASE COUNT      284 a      139 c      109 g      273 t
ORIGIN

Query Match      66.4%; Score 18.6; DB 17; Length 805;
Best Local Similarity 84.0%; Pred. No. 7.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      2 GTTCTGGGATGATCGCTCATGCTG 26
        ||| | ||||| || |||||
        Db      113 GTTAGAGGATGATGATGATG 89

Search completed: July 11, 2003, 02:26:12
Job time : 510.434 secs

```



SUMMARIES

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 259.955 Seconds

(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631B-4

Sequence: 1 **tgactcagtccttgatcatcaccttct** 30

Scoring table: IDENTITY_NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : GenEmbl : *

1: genEmb1:*
2: gb_ba:*
3: gb_hig:*
4: gb_in:*
5: gb_om:*
6: gb_ov:*
7: gb_pat:*
8: gb_ph:*
9: gb_pl:*
10: gb_pr:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_pl:*
25: em_ro:*
26: em_sy:*
27: em_un:*
28: em_vl:*
29: em_yi:*
30: em_hg_hum:*
31: em_hg_in:*
32: em_hg_other:*
33: em_hg_mu:*
34: em_hg_pl:*
35: em_hg_ro:*
36: em_hg_nam:*
37: em_hg_vrt:*
38: em_sy:*
39: em_hgo_hum:*
40: em_hgo_mu:*
41: em_hgo_other:*

Pred. No. is the number of results predicted by chance to have a

Result		Query			ID	Description
No.	Score	Match	Length	DB		
1	21.2	70.7	161776	AL133229	AL133229 Human DNA	
2	21.2	70.7	166523	AC022646	AC022646 Homo sapi	
3	21.2	70.7	198917	AL356157	AL356157 Human DNA	
4	21	70.0	6751	AF408636	AF408636 Bacterioph	
5	21	70.0	30265	AC013516	AC013516 Homo sapi	
6	21	70.0	93312	HS32110	HS32110 Human DNA	
7	21	70.0	141313	AC004943	AC004943 Homo sapi	
8	21	70.0	147091	AC106303	AC106303 Homo sapi	
9	21	70.0	151552	AC024132	AC024132 Homo sapi	
10	21	70.0	156128	AC121186	AC121186 Rattus no	
11	21	70.0	160696	AC090937	AC090937 Homo sapi	
12	21	70.0	161604	AP001978	AP001978 Homo sapi	
13	21	70.0	177875	AC025163	AC025163 Homo sapi	
14	21	70.0	182695	AC130278	AC130278 Mus muscu	
15	21	70.0	193700	AC024545	AC024545 Homo sapi	
16	21	70.0	227461	AC021429	AC021429 Homo sapi	
17	20.6	68.7	117890	AL133466	AL133466 Human DNA	
18	20.6	68.7	148851	AC016527	AC016527 Homo sapi	
19	20.6	68.7	153926	AC098093	AC098093 Rattus no	
20	20.6	68.7	160599	AC096630	AC096630 Pan trogl	
21	20.6	68.7	164435	AC112508	AC112508 Homo sapi	
22	20.6	68.7	178089	AC010929	AC010929 Homo sapi	
23	20.6	68.7	179604	AP002959	AP002959 Homo sapi	
24	20.6	68.7	185319	AL844848	AL844848 Mus muscu	
25	20.6	68.7	186111	AC096541	AC096541 Rattus no	
26	20.6	68.7	187917	AC025321	AC025321 Homo sapi	
27	20.6	68.7	193671	AC112860	AC112860 Rattus no	
28	20.6	68.7	209021	AC098878	AC098878 Mus muscu	
29	20.4	68.0	2077	AK057308	AK057308 Homo sapi	
30	20.4	68.0	152422	AC016491	AC016491 Homo sapi	
31	20.4	68.0	154433	AL732447	AL732447 Mouse DNA	
32	20.4	68.0	154936	AC021152	AC021152 Homo sapi	
33	20.4	68.0	156848	AC097710	AC097710 Homo sapi	
34	20.4	68.0	158079	AC122129	AC122129 Homo sapi	
35	20.4	68.0	160544	HS253807	AL354000 Homo sapi	
36	20.4	68.0	160918	AC069196	AC069196 Homo sapi	
37	20.4	68.0	162584	AC018594	AC018594 Homo sapi	
38	20.4	68.0	164519	AC020630	AC020630 Homo sapi	
39	20.4	68.0	168720	AC092987	AC092987 Homo sapi	
40	20.4	68.0	170814	AL844565	AL844565 Mus muscu	
41	20.4	68.0	171942	AC021115	AC021115 Homo sapi	
42	20.4	68.0	175948	AC019192	AC019192 Homo sapi	
43	20.4	68.0	176688	AC084696	AC084696 Mus muscu	
44	20.4	68.0	181899	AC015804	AC015804 Homo sapi	
45	20.4	68.0	181905	CNS01RIB	AL163052 Human chr	

ALIGNMENTS

RESULT 1	
AL133229	
LOCUS	AL133229
DEFINITION	AL133229 161776 bp DNA linear PRI 06-APR-2001
Human DNA sequence from clone RP5-1167E19 on chromosome 20 Contains	
STSs, GSSs and a Cpg island, complete sequence.	
ACCESSION	AL133229
VERSION	AL133229.40 GI:8546579
KEYWORDS	HTG, Cpg island.
SOURCE	human.
ORGANISM	Homo sapiens
	Chordata; Chordata; Vertebrata; Euteleostomi;
	Eumariota; Metazoa; Eumariota; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 161776)
AUTHORS	Bates,K.
TITLE	Direct Submission

JOURNAL
Submitted (13-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humnquay@sanger.ac.uk
COMMENT
On June 14, 2000 this sequence version replaced g1:8517286.
During difference assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EMBL, SWISSPROT, TrEMBL, WP, WORMPEP, Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20

This sequence is the entire insert of clone RP5-1167E19 The true
left end of clone RP11-179J15 is at 93503 in this sequence. The
true right end of clone RP3-50802 is at 149816 in this sequence.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP5-1167E19 is from
the library RPI-5 constructed by the group of Pieter de Jong. For
further details see
http://www.ncbi.nlm.nih.gov/bacpac/home.htm
VECTOR: pCYPAC2.

FEATURES

SOURCE

location/Qualifiers
1. 161776
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP5-1167E19"
/clone_id="RPI-5"
145..449
/note="match: STS: Em:267317"
complement(172..577)
/note="match: GSS: Em:AQ729312"
278..333
/note="28 copies 2 mer ag 76% conserved"
720..756
/note="L2 repeat: matches 2379..2415 of consensus"
1741..1785
/note="L15 copies 3 mer ata 100% conserved"
1807..1893
/note="MIR repeat: matches 39..122 of consensus"
2601..2693
/note="HAL1 repeat: matches 5..98 of consensus"
2743..2880
/note="L1R45 repeat: matches 389..525 of consensus"
3243..3403
/note="HAL1 repeat: matches 425..559 of consensus"
3404..3703
/note="ALUSg repeat: matches 1..300 of consensus"
complement(3553..4081)
/note="match: GSS: Em:B95838"
complement(3631..4052)
/note="match: GSS: Em:AQ010447"
3704..3758
/note="HAL1 repeat: matches 559..618 of consensus"
3956..4352
/note="HAL1 repeat: matches 360..749 of consensus"
4353..4643
/note="ALUSx repeat: matches 1..290 of consensus"
4644..4708
/note="HAL1 repeat: matches 295..360 of consensus"
6054..6145

repeat_region
/note="MIR repeat: matches 93..197 of consensus"
6528..6652
/note="MIR repeat: matches 47..169 of consensus"
7379..7531
/note="MIR repeat: matches 27..188 of consensus"
8058..8267
/note="MIR repeat: matches 12..262 of consensus"
8631..8827
/note="L2 repeat: matches 739..939 of consensus"
8828..9148
/note="ALU repeat: matches 1..309 of consensus"
9149..9501
/note="L2 repeat: matches 939..1277 of consensus"
9507..9559
/note="ALUB repeat: matches 249..301 of consensus"
9595..9883
/note="ALUB repeat: matches 1..252 of consensus"
9884..10018
/note="L2 repeat: matches 1262..1398 of consensus"
10076..10172
/note="MIR1 repeat: matches 104..200 of consensus"
10446..10663
/note="L1R33 repeat: matches 294..521 of consensus"
10824..10911
/note="MIR1 repeat: matches 317..410 of consensus"
10912..11493
/note="L2 repeat: matches 1396..2019 of consensus"
11505..12032
/note="L2 repeat: matches 2208..2738 of consensus"
12058..12368
/note="ALUSg repeat: matches 1..311 of consensus"
complement(12727..13254)
/note="match: GSS: Em:AQ714903"
13528..13670
/note="L2 repeat: matches 1293..1445 of consensus"
13697..13763
/note="L2 repeat: matches 2607..2674 of consensus"
14119..14269
/note="L1M4 repeat: matches 4384..4543 of consensus"
14432..14650
/note="L1M4 repeat: matches 5428..5642 of consensus"
14802..15103
/note="ALUS repeat: matches 2..303 of consensus"
15307..15725
/note="MIR1 repeat: matches 68..507 of consensus"
15745..16662
/note="L1M4 repeat: matches 3730..4663 of consensus"
16794..16875
/note="L1M4 repeat: matches 3460..3538 of consensus"
16876..17183
/note="ALU repeat: matches 1..304 of consensus"
17184..17510
/note="L1M4 repeat: matches 3115..3460 of consensus"
17439..17841
/note="match: GSS: Em:B32565"
17511..17611
/note="HAL1 repeat: matches 707..813 of consensus"
complement(17614..18091)
/note="match: GSS: Em:AQ264589"
17689..17892
/note="L1M3 repeat: matches 5524..5728 of consensus"
18401..18707
/note="ALUSg repeat: matches 1..308 of consensus"
19010..19193
/note="MIR1 repeat: matches 3..232 of consensus"
20676..20803
/note="64 copies 2 mer ag 62% conserved"
21194..21437
/note="match: GSS: Em:AQ84343"
22509..22811
/note="ALU repeat: matches 1..297 of consensus"
complement(23846..24267)
/note="match: GSS: Em:AQ562784"

```

misc_feature      complement(23854..24167)
                  /note="match: STS: Em:294595"
misc_feature      complement(23887..24265)
                  /note="match: GSS: Em:A0154014"
repeat_region     24117..24656
                  /note="LTR3 repeat: matches 21..530 of consensus"
repeat_region     24657..24891
                  /note="LTR3 repeat: matches 1..297 of consensus"
repeat_region     24894..24988
                  /note="LTR3 repeat: matches 1..297 of consensus"
repeat_region     25237..25278
                  /note="LTR3 repeat: matches 520..614 of consensus"
repeat_region     25237..25278
                  /note="LTR3 repeat: matches 16..55 of consensus"
repeat_region     25686..25696
                  /note="LTR3 repeat: matches 1..284 of consensus"
repeat_region     25974..26015
                  /note="LTR3 repeat: matches 198..238 of consensus"
repeat_region     26016..26149
                  /note="LTR3 repeat: matches 6024..6157 of consensus"
repeat_region     26150..26362
                  /note="LTR3 repeat: matches 238..456 of consensus"
repeat_region     26388..26598
                  /note="LTR3 repeat: matches 5727..5951 of consensus"
repeat_region     26860..26909
                  /note="LTR3 repeat: matches 2 mer to 92% conserved"
repeat_region     26961..27518
                  /note="LTR3 repeat: matches 489..1090 of consensus"
repeat_region     27976..28337
                  /note="LTR3 repeat: matches 3794..4167 of consensus"
repeat_region     28338..28529
                  /note="LTR3 repeat: matches 1..192 of consensus"
repeat_region     28676..28918
                  /note="LTR3 repeat: matches 4357..4593 of consensus"
repeat_region     28919..29225
                  /note="LTR3 repeat: matches 1..309 of consensus"
repeat_region     29226..29397
                  /note="LTR3 repeat: matches 4593..4759 of consensus"
repeat_region     29398..29709
                  /note="LTR3 repeat: matches 1..313 of consensus"
repeat_region     29710..30118
                  /note="LTR3 repeat: matches 4759..5157 of consensus"
repeat_region     30119..30247
                  /note="LTR3 repeat: matches 4759..5157 of consensus"

Query Match      70.7%; Score 21.2; DB 9; length 161776;
Best Local Similarity 88.5%; Pred. No. 30;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 CTCAGTCTGTGTCATCTACCTCTCT 30
    ||||| ||||| ||||| ||||| |||||
Db 76976 CTCAGCCTGTGTCATCTACCTCTCT 77001

RESULT 2
AC022646/c      AC022646 166523 bp DNA linear HTG 10-MAR-2001
DEFINITION      Homo sapiens clone RP11-28C18, WORKING DRAFT SEQUENCE, 18 unordered
                  pieces.
ACCESSION      AC022646
VERSION      AC022646.3 GI:7249102
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 166523)
Biren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-28C18
Unpublished
2 (bases 1 to 166523)
Biren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Bouhagalter,B., Brown,A., Burkett,G., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeVellano,K., Dewar,K., Domino,M., Doyle,M., Fenesstor,J.,

```

TITLE JOURNAL COMMENT

```

Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardina,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Hosland,J.C., Johnson,R., Jones,C., Kamp,L., Karatas,A., Klein,J.,
Landers,T., Lehotzky,J., Levine,R., Lien,C., Liu,S., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrum,J., Meus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Ollivier,T.M., Peterson,R.,
Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,N.
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced gi:6964422.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
Project Information
Center project name: L4741
Center clone name: 28.C.18
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 157134 bases at least Q40
Consensus quality: 161890 bases at least Q30
Consensus quality: 163469 bases at least Q20
Insert size: 156000; agarose-1p
Insert size: 164823; sum-of-coverage
Quality coverage: 5.2 in Q20 bases; agarose-1p
Quality coverage: 4.9 in Q20 bases; sum-of-coverage
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1099: contig of 1099 bp in length
* 1100 1199: gap of 100 bp
* 1200 2705: contig of 1506 bp in length
* 2706 2805: gap of 100 bp
* 2806 4225: contig of 1420 bp in length
* 4226 4325: gap of 100 bp
* 4326 5737: contig of 1412 bp in length
* 5738 5837: gap of 100 bp
* 5838 7280: contig of 1443 bp in length
* 7281 7380: gap of 100 bp
* 7381 11595: contig of 4215 bp in length
* 11596 11695: gap of 100 bp
* 11696 16125: contig of 4430 bp in length
* 16126 16225: gap of 100 bp
* 16226 21190: contig of 4965 bp in length
* 21191 21290: gap of 100 bp
* 21291 22715: contig of 6425 bp in length
* 22716 27815: gap of 100 bp
* 27816 33432: contig of 5617 bp in length
* 33433 33532: gap of 100 bp
* 33533 46593: contig of 13061 bp in length
* 46594 46693: gap of 100 bp
* 46694 58185: contig of 11492 bp in length
* 58186 58285: gap of 100 bp
* 58286 70164: contig of 11879 bp in length
* 70165 70264: gap of 100 bp

```

FEATURES
source

VERSION AL356157.14 GI:14596303
 KEYWORDS human.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE: 1 (bases 1 to 198917)
 AUTHORS Tracey, A.
 TITLE Direct Submission
 JOURNAL Submitted (03-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 COMMENT requests: clonequerry@sanger.ac.uk
 On Jul 4, 2001 this sequence version replaced GI:1420554

COMMENT

```
misc_feature 1.1059 /note="assembly-fragment"
misc_feature 1200.2705 /note="assembly-fragment"
misc_feature 2806.4425 /note="assembly-fragment"
misc_feature 4326.5737 /note="assembly-fragment"
misc_feature 5838.7280 /note="assembly-fragment"
misc_feature 7381.11593 /note="assembly-fragment"
misc_feature /note="assembly-fragment"
misc_feature 11696.16125 /note="assembly-fragment"
misc_feature 16226.21190 /note="assembly-fragment"
misc_feature 21291.27715 /note="assembly-fragment"
misc_feature 27816.33432 /note="assembly-fragment"
misc_feature 35353.46593 /note="assembly-fragment"
misc_feature 46694.58185 /note="assembly-fragment"
misc_feature 58286.70164 /note="assembly-fragment"
misc_feature clone_end:17 /note="assembly-fragment"
misc_feature 70265.84123 /note="assembly-fragment"
misc_feature 84224.100094 /note="assembly-fragment"
misc_feature 100195.116177 /note="assembly-fragment"
misc_feature 116278.134153 /note="assembly-fragment"
misc_feature 134554.166523 /note="assembly-fragment"
```

BASE COUNT	ORIGIN
49468 a	32573 c 33029 g 49750 t 1703 others

Query Match	70.7%	Score 21.2;	DB 2;	Length 166523;
Best Local Similarity	88.5%;	Pred. No. 30;		
Matches :23;	Conservative	0;	Mismatches 3;	Indels 0.

QY	4	ACTCAGTCCCTTGGTCATCTCACCTTC	29
Db	49090		49
		ACTCAGTCCCTTGGAGATCTCACCTTC	

RESULT 3	AL356157	LOCUS	AL356157	198917 bp	DNA	linear	PRI 03-JUL-2001
DEFINITION			Human DNA sequence from clone RP11-733D4 on chromosome 10, complete sequence.				
ACCESSION	AL356157						

FEATURES	Location/Qualifiers
source	1. .198917

of Placeta de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

This sequence is the entire insert of clone Rp11-733d4 The true left end of clone Rp11-523k20 is at 189938 in this sequence. The true right end of clone Rp11-236k7 is at 98701 in this sequence. The true right end of clone Rp13-37403 is at 128227 in this sequence.

```

repeat_region      289..323
repeat_region      1059..1224
misc_feature       2936..3182
repeat_region      3623..3937
repeat_region      4292..4540
repeat_region      8422..10352
repeat_region      10553..11064
repeat_region      11065..11767
repeat_region      11768..13024
repeat_region      12025..14589
misc_feature       14524..14755
/note="match: GSS: Em:A0531274 Em:A0546391 Em:A0625817"

```

```

repeat_region 14608..14681
/note="LIPB1 repeat: matches -485..-412 of consensus"
repeat_region 14732..14930
/note="LIPBa repeat: matches -1543..-1345 of consensus"
repeat_region 16415..16679
/note="L2 repeat: matches 2286..2552 of consensus"
repeat_region 16741..16845
/note="L2 repeat: matches 1038..1144 of consensus"
misc_feature 16851..17272
/note="match: GSS: Em:A0057385"
repeat_region 16937..17018
/note="L2 repeat: matches 1158..1237 of consensus"
repeat_region 17197..17666
/note="L2 repeat: matches 2169..2696 of consensus"
repeat_region 18929..18970
/note="21 copies 2 mer ac 100% conserved"
misc_feature 19952..20346
/note="match: GSS: Em:A0384380"
repeat_region 20206..20305
/note="MIR repeat: matches 29..135 of consensus"
repeat_region 20669..20770
/note="MIR repeat: matches 47..152 of consensus"
repeat_region 20958..22168
/note="LIP repeat: matches 3882..5092 of consensus"
repeat_region 22177..23069
/note="LIPa3 repeat: matches 5254..6146 of consensus"
repeat_region 23070..23109
/note="20 copies 2 mer aa 80% conserved"
repeat_region 23111..23358
/note="LIM4 repeat: matches 2275..2185 of consensus"
repeat_region 23364..23429
/note="AluJ/FRAM repeat: matches 216..281 of consensus"
repeat_region 23431..25705
/note="LIM4 repeat: matches 2215..4526 of consensus"
repeat_region 25714..25910
/note="MER1B repeat: matches 141..337 of consensus"
repeat_region 25899..25948
/note="MTRID repeat: matches 55..99 of consensus"
repeat_region 26064..26105
/note="21 copies 2 mer aa 76% conserved"
repeat_region 26107..26180
/note="MTRID repeat: matches 353..429 of consensus"
repeat_region 26179..26319
/note="MER1B repeat: matches 2..139 of consensus"
repeat_region 26323..26824
/note="LIM4 repeat: matches 4529..5009 of consensus"
repeat_region 26825..27073
/note="MTRID repeat: matches 240..505 of consensus"
repeat_region 27126..27177
/note="MTRID repeat: matches 48..99 of consensus"
repeat_region 27447..27895
/note="LIMD3 repeat: matches 5913..6401 of consensus"
repeat_region 27910..28297
/note="LIPa10 repeat: matches 5767..6159 of consensus"
repeat_region 28313..34723
/note="LIPa7 repeat: matches 2..6143 of consensus"
repeat_region 34727..35016
/note="LIPa repeat: matches 5492..5781 of consensus"
repeat_region 35017..35281
/note="Alu repeat: matches 35..294 of consensus"
repeat_region 35282..36935
/note="LIPa repeat: matches 3825..5493 of consensus"
repeat_region 36938..38320
/note="LIM3 repeat: matches 6389..7735 of consensus"
repeat_region 39478..39533
/note="MER5B repeat: matches 96..148 of consensus"
repeat_region 39756..39872
/note="MIR repeat: matches 18..140 of consensus"
repeat_region 40558..40687
/note="MIR repeat: matches 35..152 of consensus"
repeat_region 41054..41249
/note="MIR repeat: matches 24..233 of consensus"
repeat_region 41306..41596

```

```

repeat_region 41620..41731
/note="L2 repeat: matches 2571..2691 of consensus"
repeat_region 42492..42997
/note="MTRIC repeat: matches 1..464 of consensus"
repeat_region 42998..43171
/note="L2 repeat: matches 2075..2236 of consensus"
repeat_region 44118..44845
/note="LIMB3 repeat: matches 5447..6182 of consensus"
repeat_region 45984..46490
/note="LIR40a repeat: matches 2..519 of consensus"
repeat_region 47015..47093
/note="MER91A repeat: matches 105..186 of consensus"
repeat_region 47500..47917
/note="L2 repeat: matches 2309..2746 of consensus"
repeat_region 49126..49349
/note="MER20 repeat: matches 1..216 of consensus"
repeat_region 49611..49655
/note="MIR repeat: matches 219..262 of consensus"
repeat_region 49662..50194
/note="L2 repeat: matches 2096..2661 of consensus"
repeat_region 51919..52224
/note="AluSg repeat: matches 1..306 of consensus"
repeat_region 53562..53818
/note="LIMB8 repeat: matches 5819..6167 of consensus"
repeat_region 53817..54341
/note="LIMB6 repeat: matches 5614..6140 of consensus"
repeat_region 54343..54633
/note="AluSg repeat: matches 1..291 of consensus"
repeat_region 56076..56128
/note="L2 repeat: matches 2662..2709 of consensus"
repeat_region 56706..56750
/note="match: GSS: Em:A0146681"
repeat_region 58317..58635
/note="match: GSS: Em:A0146681"
repeat_region 58311..58743
/note="match: GSS: Em:A0890399"
repeat_region 58927..59200
/note="L2 repeat: matches 1543..1831 of consensus"
repeat_region 59422..59806
/note="LIM4 repeat: matches 472..862 of consensus"
repeat_region 59871..60217
/note="LIM4 repeat: matches 867..1199 of consensus"
repeat_region 60292..60367
/note="LIM4 repeat: matches 1204..1281 of consensus"

```

```

Query Match 70.7% Score 21.2 DB 9 Length 198917;
Best Local Similarity 88.5% Pred. No. 30;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Cy 4 ACGAGTCCTTGGTCATCTGCCTTC 29
Db 187890 ACTGAGTCCTTGGATCTCCTTC 187915

RESULT 4
AF408636
LOCUS
DEFINITION Bacteriophage phi-12 P14 (14), P15 (15), core protein P7 (7),
RNA-dependent RNA polymerase P2 (2), NTPase P4 (4), and major
structural protein P1 (1) genes, complete cds.
ACCESSION AF408636
VERSION AF408636
KEYWORDS AF408636.1 GI:15488102
SOURCE bacteriophage phi-12.
ORGANISM bacteriophage phi-12.
REFERENCE 1 (bases 1 to 6751)

```

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE

Gottlieb, P.J., Potgieter, C., Wei, H. and Toporovsky, I.
Characterization of Bacteriophage phi12
Unpublished
2 (bases 1 to 6751)
Gottlieb, P.J., Potgieter, C., Wei, H. and Toporovsky, I.
Direct Submission
Submitted (10-AUG-2001) Microbiology and Immunology, Sophie Davis
School of Biomedical Education, City College of New York, 138th
Street and Convent Avenue, New York, NY 10031, USA
Location/Qualifiers
1. 6751
/organism="bacteriophage phi-12"
/db_xref="taxon:161736"
410. 709
/gene="14"
410. 709
/gene="14"
/codon_start=1
/transl_table=11
/product="p14"
/protein_id="AA01104.1"
/db_xref="GI:15488103"
/translation="MRKGRKMNOKPDLITTSVNVNRRNTNRRVANSKOYWECCG
PALNLEFKKCVFYGSAIRLMLVGLDSKTEDEHVALEEMGYNEGSFEIVKE"
714. 1016
/gene="15"
714. 1016
/gene="15"
/codon_start=1
/transl_table=11
/product="p15"
/protein_id="AA01105.1"
/db_xref="GI:15488104"
/translation="WFGSDNGAVHTOTTINPVKWKSGSPSEKSTVLYLFTRGA
IRGRDIPGKAGHAILLCLMDGSVFQMMIDHAPRGKTIETIRPDSTSTHYILG"
1019. 1528
/gene="7"
1019. 1528
/gene="7"
/codon_start=1
/transl_table=11
/product="core protein p7"
/protein_id="AA01106.1"
/db_xref="GI:15488105"
/translation="MDFTDMSKNORLEONRLAJOYETSLMWSHNGDVPVTTGNNV
ELGALHEVSWRVLLNDIADQLKAMTSHLELSTKNOEQARAKDGTVPVDFD
HGAIDDADE"
1512. 3491
/gene="2"
1512. 3491
/gene="2"
/codon_start=1
/transl_table=11
/product="RNA-dependent RNA polymerase p2"
/protein_id="AA01107.1"
/db_xref="GI:15488106"
/translation="MMPINSEPKHAETIIVDAPRNSVGDHLENGCVNHDPPDA
YTKMGPIAGVDPBNPTLIFANAROTGOLVAPKARILGIGTSFEPTVAFDREY
KRAMVAOPDLNDSGLNGLHTFDSIRSGYQWYVSLPNSVGLLDLLOER
MTAROVVIRREIKRVYMGHMKPTAIKIKRSTSGPPRNVNAEKILYALAFSGNRY
NGYLAFSGDLSPRYDYEAAYVINGIVNRQVONPGKRRVNAQADIERLAVSKRP
ITTKVINGVYDDPAARNTRIYVAGPWTIVVAQPFATGCMAMPFELYSRATWHPDD
KTAGLEKHAFFGVSSYDHSFSEKIDLSLEVEKEISEIEMELASSI.FYAAVPR
PLGPDGPOLVGNPNRILEKQVKNRSGHAFSLPAKVMVVIDTVSRKPKMGIDVVA
NNDALIKGDMPGCINNDDDEIVWFKSRSDVRLRLILLETQPORORFKGPESGAVF
SGSVOLGIPAKYAVERTITPPORITCPEKSTGCRKRPVGLIERKRNKSHVPL
SEWVVFDDTYATILAEPIHYSGLIGVORAKHEIFSYDDLSMKEIIVLDPNKMTHRF
TDEERDVOSAFRKRQIPTEFEKMKHNGTV"
3488. 4483
/gene="4"
3488. 4483
/gene="4"

/codon_start=1
/transl_table=11
/product="NPase p4"
/protein_id="AA01108.1"
/db_xref="GI:15488107"
/translation="MHTIDAKSPAKIRAAQVAFTDAPGSWFHDFTSGVLSVEDGT
PVLAIGVSGDAIEVDNAQRIVAYREKSVKADGSSVYVYNGEMKQGHGMVLDL
TGEIVGSPVAAYFEFGHRSVAGVAVYKSKNSGKTPLVHAGLGGKRYANVYRGE
PLSGVNTDFNVFVDDIARAMLOHRVAVIYDLSKLVICAGAGNTSGISGADLSDI
GAMASGCVAVIASINPTSNDDIIVELVYKASNSSTSLYSTDDGEMQVLRKRGEG
4497. 6668
/gene="11"
4497. 6668
/gene="11"
/codon_start=1
/transl_table=11
/product="major core structural protein p1"
/protein_id="AA01109.1"
/db_xref="GI:15488108"
/translation="MAKKRIDKQAQANRVRPERPNTSMNSSLATAMANSVRIENT
VDRASSTVEEVCSEVNLVVAHYIKEPGRVRENVAKPFLHRRVYVTEANTSNV
STGEAQTLDGLRSLNSAIEDAALICEVPIPLAKGLDRIKTRTSIEFPK
TVSIALDLSIDLAQNAVRVAETVNVRMHSEKMSYLAARFGLVAVIGELHRSV
VOHIFDDIVKALHVKLIAASADMGVSDSRILHHPVIDESRNTVTSALDMPDIG
DGLKNDIHTLEKEAPIALASLKSGRVNVGRDRLTYGKKTIVNVEGEPFIFIG
RNASLASVACQSVPNDAIMFROAKNVVPGDITKFEVASSMPKNSGVGFVNEIT
NATILHAGCGRSTYNEGSAVAKYIGEAFLPGDDGENSEKELVLMADACVANT
GPEKVENYISTHETRYNEDGSAVAKYIGEAFLPGDDGENSEKELVLMADACVANT
ACOLSDKALFTRVLDLEDKRTVEKEREVFNKGAETISGAVVTEGMOGLPDER
FVYPVHNLVMDITIAIHSSFNELINAKKVASRPPGEGDGLDVSNTVYVQLVSF
NMLSRRAILTFEASQVPPQYROIYAVSARMAVANKASDILRLRGTMOQSSFNAYAD
VGLVILATITAGLSKFRDLLEASVNEALMLTGSRYVAN"

BASE COUNT 1612 a 1801 c 1852 g 1486 t

Query Match 70.0%; Score 21; DB 7; Length 6751;
Best local similarity 82.8%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGGACTCAGTCTTGCATCTCACCCTTC 29
DB 4449 TGGATCGCTCTTCGTCTCTACCTCC 4477

RESULT 5
AC013516/c
LOCUS
DEFINITION Homo sapiens clone RP11-115K5, LOW-PASS SEQUENCE SAMPLING.
AC013516
VERSION AC013516.2 GI:9112504
KEYWORDS HMG; HMGs, PHASED.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 30265)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-115K5
Unpublished
2 (bases 1 to 30265)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
Cooke, P., Deatellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., Fitzhugh, M., Forrest, C., Funk, R., Gage, D.,
Galagan, J., Gardy, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howard, J., Johnson, R., Jones, C., Kamm, B., Karatas, A., Klein, J.,
Lehoczky, D., Lien, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McKernan, K., McKernan, K., McDonald, P., Marquis, N.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

TITLE JOURNAL COMMENT

Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6403765.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: 115_K_5

* NOTE: This record contains 34 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 753: contig of 753 bp in length

754 853: gap of 100 bp

854 1682: contig of 829 bp in length

1683 1782: gap of 100 bp

2620: contig of 838 bp in length

2621 2720: gap of 100 bp

2721 3546: contig of 826 bp in length

3547 3646: gap of 100 bp

3647 4551: contig of 905 bp in length

4552 4651: gap of 100 bp

4652 5461: contig of 810 bp in length

5462 5561: gap of 100 bp

5562 6319: contig of 758 bp in length

6320 6419: gap of 100 bp

7229: contig of 810 bp in length

7230 7329: gap of 100 bp

7330 8105: contig of 776 bp in length

8106 8205: gap of 100 bp

8206 9031: contig of 826 bp in length

9032 9131: gap of 100 bp

9132 9895: contig of 764 bp in length

9896 9995: gap of 100 bp

10791 10890: contig of 795 bp in length

10891 10890: gap of 100 bp

10891 11726: contig of 836 bp in length

11727 11826: gap of 100 bp

11827 12605: contig of 779 bp in length

12606 12705: gap of 100 bp

12706 13525: contig of 820 bp in length

13526 13625: gap of 100 bp

13626 14374: contig of 749 bp in length

14375 14474: gap of 100 bp

14475 15188: contig of 714 bp in length

15189 15288: gap of 100 bp

15289 16131: contig of 843 bp in length

16132 16231: gap of 100 bp

16232 16933: contig of 702 bp in length

16934 17033: gap of 100 bp

17034 17842: contig of 809 bp in length

17843 17942: gap of 100 bp

17943 18731: contig of 789 bp in length

18732 18831: gap of 100 bp

18832 19608: contig of 777 bp in length

19609 19708: gap of 100 bp

FEATURES

SOURCE

19709 20487: contig of 779 bp in length
* 20488 20587: gap of 100 bp
* 20588 21395: contig of 808 bp in length
* 21396 21495: gap of 100 bp
* 21496 22261: contig of 766 bp in length
* 22262 22361: gap of 100 bp
* 22362 23158: contig of 793 bp in length
* 23159 23258: gap of 100 bp
* 23259 24052: contig of 794 bp in length
* 24053 24152: gap of 100 bp
* 24153 24927: contig of 775 bp in length
* 24928 25027: gap of 100 bp
* 25028 25812: contig of 780 bp in length
* 25813 25912: gap of 100 bp
* 25913 26717: contig of 805 bp in length
* 26718 26817: gap of 100 bp
* 26818 27628: contig of 811 bp in length
* 27629 27728: gap of 100 bp
* 27729 28468: contig of 740 bp in length
* 28469 28568: gap of 100 bp
* 28569 29356: contig of 788 bp in length
* 29357 29456: gap of 100 bp
* 29457 30265: contig of 809 bp in length.
Location/Qualifiers
1. 30265
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Rpl1-115K5"
/clone_11b="RRC1-11 Human Male BAC"

BASE COUNT 7067 a 6735 c 6224 g 6406 t 3833 others
ORIGIN

Query Match 70.0% Score 21; DB 2; Length 30265;
Best Local Similarity 82.8% Pred No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGACACAGTCCTGGTCATCCACCTTC 30
|||||
DB 327 GGACACAGTCCTGGTCATCCACCTTC 299

RESULT 6 93312 bp DNA linear PRI 12-DEC-1999
HS32110
LOCUS
DEFINITION Human DNA sequence from clone RPI-32110 on chromosome 22 Contains a
pseudogene similar to TCEB1L (transcription elongation factor B
(SIL1), polypeptide 1-like), a CA repeat (D22S1159), STRs and
GSSs n, complete sequence.

ACCESSION

285994 GI:2342582

VERSION

HTG: D22S1159; repeat polymorphism; TCEB1L.

SOURCE

Homo sapiens.

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

1 (bases 1 to 93312)

AUTHORS

Ho, S.

TITLE

Direct Submission

JOURNAL

COMMENT

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquary@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Aug 24, 1997 this sequence version replaced gi:1841910.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence has been finished according to sequence map criteria
as follows: An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'

Feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/Map/Chr22> RPI-32110 is from the library RPI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR> PCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RPI-32110 it may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone RPI-32110 is at 1 in this sequence.

FEATURES

SOURCE

```

1. 93312
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="22"
   /clone_1lb="RPI-1"
   25..196
misc_feature
   /note="match: STS: Em:HS32110T"
   671..810
repeat_region
   /note="35 copies 4 mer ggtg 70 conserved"
   1041..1091
repeat_region
   /note="L2 repeat: matches 2688..2746 of consensus"
   1051..1126
repeat_region
   /note="MIR repeat: matches 174..256 of consensus"
   1454..1608
repeat_region
   /note="L2 repeat: matches 1986..2144 of consensus"
   1609..1904
repeat_region
   /note="AluSg repeat: matches 1..308 of consensus"
   1905..1959
repeat_region
   /note="L2 repeat: matches 2144..2197 of consensus"
   2224..2389
repeat_region
   /note="MIR repeat: matches 89..262 of consensus"
   2471..2550
repeat_region
   /note="L2 repeat: matches 2667..2745 of consensus"
   3038..3106
repeat_region
   /note="MIR repeat: matches 56..128 of consensus"
   3294..3530
repeat_region
   /note="MIR repeat: matches 13..258 of consensus"
   4209..4325
repeat_region
   /note="MIR repeat: matches 48..175 of consensus"
   6405..6487
repeat_region
   /note="MIR repeat: matches 46..131 of consensus"
   6608..6910
repeat_region
   /note="AluSg repeat: matches 1..303 of consensus"
   8780..8899
repeat_region
   /note="L2 repeat: matches 2091..2218 of consensus"
   8908..8971
repeat_region
   /note="L2 repeat: matches 2388..2453 of consensus"
   8972..9266
repeat_region
   /note="AluX repeat: matches 1..296 of consensus"
   9267..9581
repeat_region
   /note="L2 repeat: matches 2453..2750 of consensus"
   complement(9882..10117)
misc_feature
   /note="match: STS: Em:Z53151; match: STS: Em:HSB0292H9"
   9936..9977
repeat_region
   /note="21 copies 2 mer tg 95 conserved"
   9939..9978
repeat_region
   /note="10 copies 4 mer ggtg 95 conserved"
   10003..10283
repeat_region
   /note="AluSg repeat: matches 1..280 of consensus"
   10524..10766
repeat_region
   /note="L2 repeat: matches 2260..2498 of consensus"
   11167..11346

```

```

   /note="MIR repeat: matches 1..184 of consensus"
   11347..11661
repeat_region
   /note="AluX repeat: matches 3..306 of consensus"
   11662..11892
repeat_region
   /note="MIR repeat: matches 184..400 of consensus"
   11893..12203
repeat_region
   /note="AluX repeat: matches 1..310 of consensus"
   12204..12256
repeat_region
   /note="MIR repeat: matches 400..453 of consensus"
   12368..12523
repeat_region
   /note="MIR repeat: matches 108..262 of consensus"
   12810..12915
repeat_region
   /note="L2 repeat: matches 2438..2536 of consensus"
   12915..12987
repeat_region
   /note="MIR repeat: matches 56..121 of consensus"
   13585..13632
repeat_region
   /note="L2 repeat: matches 2649..2698 of consensus"
   13826..14242
repeat_region
   /note="MIR repeat: matches 1..466 of consensus"
   14858..14980
repeat_region
   /note="MIR repeat: matches 56..192 of consensus"
   15124..15405
repeat_region
   /note="AluSg repeat: matches 1..288 of consensus"
   15489..15585
repeat_region
   /note="MIR repeat: matches 59..148 of consensus"
   15637..15699
repeat_region
   /note="L2 repeat: matches 2348..2411 of consensus"
   16503..16575
repeat_region
   /note="MIR repeat: matches 27..104 of consensus"
   16708..17082
repeat_region
   /note="MIR repeat: matches 1..382 of consensus"
   18010..18204
repeat_region
   /note="MIR repeat: matches 35..239 of consensus"
   18260..18357
repeat_region
   /note="MIR repeat: matches 42..144 of consensus"
   19043..19101
repeat_region
   /note="MIR repeat: matches 92..152 of consensus"
   19587..19718
repeat_region
   /note="MIR repeat: matches 25..185 of consensus"
   20352..20528
repeat_region
   /note="MIR repeat: matches 33..212 of consensus"
   20896..21504
repeat_region
   /note="L2 repeat: matches 3..610 of consensus"
   21540..21702
repeat_region
   /note="MIR repeat: matches 61..233 of consensus"
   22247..22605
repeat_region
   /note="MIR repeat: matches 193..547 of consensus"
   23244..23555
repeat_region
   /note="AluSg repeat: matches 1..310 of consensus"
   23729..23843
repeat_region
   /note="MIR repeat: matches 34..144 of consensus"
   24584..24940
repeat_region
   /note="L2 repeat: matches 2315..2671 of consensus"
   25020..25155
repeat_region
   /note="MIR repeat: matches 60..188 of consensus"
   25342..25692
repeat_region
   /note="L2 repeat: matches 2124..2478 of consensus"
   26105..26211
repeat_region
   /note="MIR repeat: matches 53..160 of consensus"
   26212..26526
repeat_region
   /note="AluX repeat: matches 1..308 of consensus"
   26527..26832
repeat_region
   /note="MIR repeat: matches 160..459 of consensus"
   27434..27544
repeat_region
   /note="MIR repeat: matches 80..195 of consensus"
   28231..28542
repeat_region
   /note="AluX repeat: matches 1..312 of consensus"
   30366..31108
repeat_region
   /note="L2 repeat: matches 1027..1841 of consensus"
   31858..33268
repeat_region
   /note="L2 repeat: matches 2210..3688 of consensus"
   33136..33471
misc_feature
   /note="match: GSS: Em:AG617050"

```



```

repeat_region      33470..33545
                    /note="19 copies 4 mer cct 80 conserved"
repeat_region      33547..33682
                    /note="68 copies 2 mer ct 73 conserved"
repeat_region      33686..33750
                    /note="MIR1D repeat: matches 182..249 of consensus"
repeat_region      34043..35036
                    /note="MIR4 repeat: matches 3723..4744 of consensus"
repeat_region      35209..35381
                    /note="MIR repeat: matches 23..219 of consensus"
repeat_region      36094..36230
                    /note="MIR repeat: matches 117..260 of consensus"
repeat_region      36714..36893
                    /note="MIR repeat: matches 79..262 of consensus"
misc_feature       37498..37754
                    /note="match: STS: Em:G49322"
misc_feature       37579..37934
                    /note="match: STS: Em:G49323"
repeat_region      37971..38036
                    /note="L2 repeat: matches 2127..2194 of consensus"
repeat_region      38072..38180
                    /note="MIR repeat: matches 16..143 of consensus"
repeat_region      38333..38579
                    /note="L2 repeat: matches 2358..2671 of consensus"

Query Match      70.0%; Score 21; DB 9; Length 93312;
Best Local Similarity 82.8%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 TGGACTCAGCTCTGGTCATCTCACCTTC 29
    ||| ||| ||| ||| ||| ||| ||| |||
Db 14888 TGAAGTCAAGCCCTGCTCACTACCTTC 14916

RESULT 7
AC004943/C      141313 bp      DNA      linear      PRI 22-MAY-2002
LOCUS      AC004943
DEFINITION      Homo sapiens PAC clone RP5-991G20 from 16, complete sequence.
ACCESSION      AC004943
VERSION      AC004943.2 GI:21070767
KEYWORDS
SOURCE      HTG.
ORGANISM      Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 141313)
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
9847074
2 (bases 1 to 141313)
Kozlowicz,A., Lapiant,Y. and Hottic,M.
The sequence of Homo sapiens PAC clone RP5-991G20
Unpublished (2001)
3 (bases 1 to 141313)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 141313)
Waterston,R.
Direct Submission
Submitted (24-NOV-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 141313)
Waterston,R.
Direct Submission
Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 141313)
Waterston,R.

```

```

TITLE      Direct Submission
JOURNAL    Submitted (22-MAY-2002) Department of Genetics, Washington
COMMENT    University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           On May 22, 2002 this sequence version replaced gi:3924671.

           ----- Genome Center
           Center: Washington University Genome Sequencing Center
           Center code: WUGSC
           Web site: http://genome.wustl.edu/gsc
           Contact: sapiens@wustl.wustl.edu
           ----- Summary Statistics
           Center project name: R.DJ091620

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPI-5, prepared by
Pleier de Jong and coworkers at http://www.chori.org using the
method described by Ioannou et al., Nature Genetics 6:84-9 (1994).
The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc.
(http://www.genomesystems.com) or Research Genetics, Inc.
(http://www.resgen.com); or from Pleier de Jong.
VECTOR:      pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is CTD-2532L16, 200 bp overlap.
Actual start of this clone is at base position 1 of RP5-991G20;
actual end is at base position 141313 of RP5-991G20.
Location/Qualifiers
1. 141313
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="16"
   /map="16"
   /clone="RP5-991G20"
   /clone_11b="RPI-5"
85..235
   /rpt_family="MIR"
1219..1292
   /rpt_family="L2"
1386..1786
   /note="similar to EST W87623 (NID:91401687) zh65h07.r1"
2123..2228
   /rpt_family="L2"
2352..2613
   /rpt_family="Alu"
2743..2916
   /rpt_family="L1"
2978..3259
   /rpt_family="Alu"
3396..3517
   /rpt_family="Alu"
3521..4092
   /rpt_family="L1"
4285..4342
repeat_region

```

REBDSABEGGNTGKETLTPDALKAKKKEMLPAGGSESSMLPPLASDARGNAPFKAL
LENFEBCELEIDYNNENKOROKNGKCTDGOENLECEBSCGLTFNNILILSHOETHY
QWYFFPELOLEFFAIDYNDHYDKYLPCTDGOENLECEBSCGLTFNNILILSHOETHY
PASAPITSPILASAOSSVLPLOLSMDELPISPLMOTNPLOLPLAQDPPQAPV
PLPAIALOYHOLANOSVLPLOLONOKRPTRTITDLOLVKLOFIDINNSSEPOLEIEMAD
KSGLEOKYIKHMPENTFKRQKNNKSPNPNFNPSTLEBLKIDRSPSEPEKQY
MGRSSSTRTPTDQLAVLDPFDPAAYAKRDEBFLNLTIPRVLVWFOAAROK
ARKTENEDEBESREELINDNYITSMILNOKCKSLIVORFIDILKHOIKCTYD
EDELBOODSONESMDAMELITPASPCCPTMPOSSOYASNPANNSAFTLOLRE
ABELETANFSTKEGDEKPLKAPSSQCTPMSOCTPMSOCTPMSOCTPMSOCTPMSOCT
LPOLVSLPPLPPAPAPPOCLOPSSPSSPOLSHLPKLPKLPSTPOQLANPLP
PYCCTDGOENLECEBSCGLTFNNILILSHOETHYQWYFFPELOLEFFAIDYNDHYDKYLP
LITLGAITPOTILPASANSPTSPITSMATLKRKILKEEASSPENDSGGCEPODRIL
RATTPOLITILOKLULFANSPKMLTHIAHEMLKRVVYUWFOATRAREKGOFR
AYGPAOAHARCPCLRALFKAKLALNIASTIRHIMBAKRGVNLMSMLDCCGLO
KAGFDISTSTSPSSSGSOVNLIPVSKYENLEPRLISSPTIKYEGIDEPESB
SSVNLINDOKRLNDCSSVNLIDTPITDGBGANDNANSGTILKHOIKCTYD

Query Match	70.0%	Score 21;	DB 9;	Length 141313;
Best Local Similarity	82.8%	Prod No	37	

best local similarity 82.8%; pred. no. 37;
Matches 24; conservative 0; Mismatches

0; Gaps 0;

xy 2 GGACTCAGTCCTGGTCATCTCACCTTCT 30
|||||
Db 77445 GGACTCAGTCCTAGGTACGCTCCCAAAAGC 77

RESULT 8
AC108030/c

DEFINITION	147091 bp	DNA	linear	PRI 20-MAR-2002
ACCESSION	AC108030			
DESCRIPTION	Homo sapiens BAC clone RP11-100C9 from 4, complete sequence.			
ACCESSION	AC108030	AC000000		

AC108030.4 GI:19551171

KEYWORDS	HTG.
SOURCE	HTG.

DOORCE
ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL
OF THE
MEDICAL SOCIETY

PUBMED

REFERENCE

AUTHORS
TITLE

JOURNAL

REFERENCE

AUTHORS
TITLE

JOURNAL,

REFERENCE

AUTHORS

JOURNAL OF THE

CONCLUSIONS

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

REFERENCE

10-10-68

JOURNAL,

COMMENT

1

Drafting Center: MIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catalanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-384K6; the clone sequenced to the right is RP11-358D17, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-100C9; actual end is at base position 34330 of RP11-358D17.

The sequence of AC009865 has been incorporated into AC108030.

FEATURES

```

source
    Location/Qualifiers
        1..147091
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="4"
        /map="4"
        /clone="RP11-100C9"
        /clone_lib="RPCI-11"
        76..372
        /rpt_family="Alu"
repeat_region
    /note="match to EST B1048881 (NID:g14455503)"
misc_feature
    1785..2022
    /note="match to EST B1048881 (NID:g14455503)"
misc_feature
    2787..2814
    /note="match to EST AM805082 (NID:g7856965)"
repeat_region
    3137..3197
    /rpt_family="AT-rich"
repeat_region
    3269..3554
    /rpt_family="Alu"
misc_feature
    4079..4181
    /note="similar to EST BM461794 (NID:g18510834)"
misc_feature
    4082..4181
    /note="match to EST AU130547 (NID:g10990901)"
misc_feature
    4082..4181
    /note="match to EST AU130547 (NID:g10990901)"
misc_feature
    4788..4980
    /note="match to EST BG687120 (NID:g13918517)"
misc_feature
    4788..4980
    /note="similar to Mus musculus EST BF302831 (NID:g11249388)"
misc_feature
    4789..4980
    /note="match to EST AU130547 (NID:g10990901)"
misc_feature
    4789..4980
    /note="match to EST BG687120 (NID:g13918517)"
misc_feature
    4789..4980
    /note="similar to EST BM461794 (NID:g18510834)"
repeat_region
    6760..6907

```

```

misc_feature
    9813..9937
    /rpt_family="MER1_type"
misc_feature
    9813..9937
    /note="match to EST AM805082 (NID:g7856965)"
misc_feature
    9813..9937
    /note="similar to EST BM461794 (NID:g18510834)"
misc_feature
    9813..9937
    /note="similar to Mus musculus EST BF302831 (NID:g11249388)"
misc_feature
    9813..9886
    /note="match to EST BG687120 (NID:g13918517)"
misc_feature
    9813..9834
    /note="match to EST AU130547 (NID:g10990901)"
repeat_region
    10777..10816
    /rpt_family="AT-rich"
repeat_region
    11253..11548
    /rpt_family="L2"
repeat_region
    12246..12801
    /rpt_family="L1"
misc_feature
    12483..12711
    /note="match to EST AM805082 (NID:g7856965)"
repeat_region
    12802..13108
    /rpt_family="Alu"
repeat_region
    13109..13147
    /rpt_family="L1"
repeat_region
    13299..13356
    /rpt_family="L2"
repeat_region
    13357..13390
    /rpt_family="TGAA)n"
misc_feature
    13489..13687
    /note="similar to Homo sapiens EST B1048935 (NID:g14455557)"
misc_feature
    13497..13826
    /note="match to EST BF910439 (NID:g12301897)"
repeat_region
    13659..13745
    /rpt_family="MER1_type"
repeat_region
    13915..14163
    /rpt_family="L2"
repeat_region
    14366..14604
    /rpt_family="MIR"
repeat_region
    16289..16314
    /rpt_family="AT-rich"
misc_feature
    16878..16879
    /note="match to EST B1712490 (NID:g15688185) le08e05.x1"
misc_feature
    16889..17009
    /note="match to EST BF031310 (NID:g10739022)"
misc_feature
    16894..17009
    /note="similar to EST BM461794 (NID:g18510834)"
misc_feature
    16894..17009
    /note="similar to Mus musculus EST BF302831 (NID:g11249388)"
misc_feature
    17068..17577
    /note="match to EST B1712490 (NID:g15688185) le08e05.x1"
misc_feature
    17411..17934
    /note="match to EST B1712728 (NID:g15688423) le08e05.y1"
misc_feature
    18993..19340
    /note="match to EST BF841853 (NID:g12195175)"
misc_feature
    20265..20409
    /note="similar to Mus musculus EST BF302831 (NID:g11249388)"
misc_feature
    20265..20403
    /note="match to EST BF031310 (NID:g10739022)"
misc_feature
    20265..20403
    /note="match to EST BM461794 (NID:g18510834)"
misc_feature
    20274..20403
    /note="match to EST BE247482 (NID:g9099239)"
misc_feature
    20311..20403
    /note="match to EST AM96773 (NID:g8257007)"
misc_feature
    20316..20403
    /note="match to EST AM369711 (NID:g6874365)"
misc_feature
    20350..20403
    /note="match to EST BE831699 (NID:g10264077)"
repeat_region
    20487..20786
    /rpt_family="Alu"

```

```

misc_feature      21804..22199
                  /note="match to EST AV707981 (MID:g10725246)"
repeat_region     21946..22011
                  /rpt_family="MIR"
repeat_region     26140..26182
                  /rpt_family="AT-rich"
repeat_region     26650..27046
                  /rpt_family="AT-rich"
misc_feature      26984..27338
                  /note="match to EST BF059079 (MID:g10812975)"

Query Match      70.0% Score 21: DB 9; Length 147091;
Best Local Similarity 82.8% Pred. No. 37;
Matches 124; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGACTCAGTCCTTGTCATCTCACTCTCT 30
    ||||||| ||||||| ||||||| |||||
Db 51688 GGCTCAGTACTTGTCATCTCTCTCT 51660

RESULT 9
AC024132 151552 bp DNA linear PRI 07-NOV-2001
LOCUS Homo sapiens BAC clone RP11-415C15 from 4, complete sequence.
ACCESSION AC024132
VERSION AC024132.7 GI:14589728
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 151552)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 98063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 151552)
AUTHORS Goya,E. and Abbott,A.
TITLE The sequence of Homo sapiens BAC clone RP11-415C15
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 151552)
TITLE Waterston,R.H.
JOURNAL Direct Submission
AUTHORS Submitted (24-FEB-2000) Genome Sequencing Center, Washington
TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63108, USA
REFERENCE 4 (bases 1 to 151552)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Genome Sequencing Center, Washington
TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63108, USA
REFERENCE 5 (bases 1 to 151552)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
TITLE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL On Jul 3, 2001 this sequence version replaced g1:14151042.
COMMENT Center: Washington University Genome Sequencing Center
          Center code: WUGSC
          Web site: http://genome.wustl.edu/gsc
          Contact: saplens@wustl.wustl.edu
          Summary Statistics
          Center project name: H_NH0415C15

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPI1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateo,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is AC006568. Actual start of this clone is at base position 1 of RP11-415C15.

The region between 36765 to 36883 is covered only by a single plasmid subclone.

FEATURES

```

source
location/Qualifiers
1..151552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-415C15"
/clone_id="RPI1-11"
2..947
/rpt_family="L1"
988..1084
/rpt_family="L1"
2351..2559
/rpt_family="MALR"
2745..3273
/rpt_family="MERL-type"
3625..3931
/rpt_family="Alu"
4992..5248
/rpt_family="Alu"
5470..5776
/rpt_family="Alu"
5629..5658
/rpt_family="(T)n"
6350..6682
/rpt_family="ERV1"
6999..7352
/rpt_family="MALR"
7350..7379
/rpt_family="AT-rich"
7353..7660
/rpt_family="Alu"
7661..7769
/rpt_family="MALR"
7853..7878
/rpt_family="(T)n"
8615..8645
/rpt_family="(TTTC)n"
8619..8929
/rpt_family="Alu"
9339..9651
/rpt_family="Alu"
9627..9651
repeat_region

```

```

repeat_region      /rpt_family="(A)n"
10012..10093
repeat_region      /rpt_family="MERL_type"
11138..11326
repeat_region      /rpt_family="MERL_type"
11397..11758
repeat_region      /rpt_family="MALR"
12010..12337
repeat_region      /rpt_family="Alu"
12315..12337
repeat_region      /rpt_family="(GAAA)n"
12522..12700
repeat_region      /rpt_family="CR1"
13352..13648
repeat_region      /rpt_family="Alu"
13661..13790
repeat_region      /rpt_family="MALR"
13791..13972
repeat_region      /rpt_family="MERL_type"
13973..14064
repeat_region      /rpt_family="MALR"
14300..14613
repeat_region      /rpt_family="MIR"
15241..16178
repeat_region      /rpt_family="L1"
16173..16901
repeat_region      /rpt_family="L1"
17238..17666
repeat_region      /rpt_family="L1"
18792..18885
repeat_region      /rpt_family="L2"
19121..19301
repeat_region      /rpt_family="L2"
19294..19319
repeat_region      /rpt_family="(TGA)n"
19954..20186
repeat_region      /rpt_family="ERVU"
21116..21231
repeat_region      /rpt_family="MALR"
21289..21485
repeat_region      /rpt_family="CR1"
22221..22536
repeat_region      /rpt_family="Alu"
22509..22536
repeat_region      /rpt_family="(A)n"
22595..22802
repeat_region      /rpt_family="L1"
22830..22940
repeat_region      /rpt_family="L1"
22941..23305
repeat_region      /rpt_family="MALR"
23306..23393
repeat_region      /rpt_family="L1"
23407..23698
repeat_region      /rpt_family="Alu"
23700..23809
repeat_region      /rpt_family="(TA)n"
23813..24012
repeat_region      /rpt_family="L1"
24045..24386
repeat_region      /rpt_family="L2"
24412..24522
repeat_region      /rpt_family="ERVU"
24602..24728
repeat_region      /rpt_family="ERVU"
26104..26128
repeat_region      /rpt_family="AT-rich"
26508..26576
repeat_region      /rpt_family="MALR"
26712..27121
repeat_region      /rpt_family="MALR"
27122..27419
repeat_region      /rpt_family="Alu"

```

```

repeat_region      27420..27680
repeat_region      /rpt_family="MALR"
27682..27951

Query Match
Best Local Similarity 82.8%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      2 GGACTGACCTCTGTGTCATCTCACCTTCT 30
Db      112885 GGGCTGACCTCTGTGTCATCTCTTCTCT 112913

RESULT 10
AC121186/c
LOCUS      156128 bp DNA linear HTG 23-JUL-2002
DEFINITION Rattus norvegicus clone CH230-331A2, *** SEQUENCING IN PROGRESS
ACCESSION AC121186
VERSION AC121186.2 GI:21902677
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 156128)
Muzny,D.M., Adams,C., Adlo-Oduola,B., All-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbata,J., Benton,J., Blumage,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.J.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,U., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Dayila,M.L., Davys,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthett,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhardt,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,U., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,D.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., McWhiney,E., McLeod,M.P., Meador,M., Mei,G., Metkay,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,S.,
Nguyen,N., Nickerson,E., Nwokoko,S., Ogulu,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Swatek,A., Tabor,P., Tameris,A., Tameris,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usman,I., Vasquez,L., Vera,V., Villalobos,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Woodden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 156128)
REFERENCE Worley,K.C.
JOURNAL Direct Submission
TITLE Submitted (16-MAY-2002) Human Genome Sequencing Center, Department

```

REFERENCE
AUTHORS
TITLE
JOURNAL

Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
(bases 1 to 156128)

Submitted (23-JUL-2002)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jul 18, 2002 this sequence version replaced g1:20806215.

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GMP

Center clone name: CH230-331A2

Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap: version 0.990329

Consensus quality: 104035 bases at least Q40

Consensus quality: 110796 bases at least Q30

Consensus quality: 115903 bases at least Q20

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 57 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

```

1      1029: contig of 1029 bp in length
1030   1129: gap of unknown length
1130   2199: contig of 1070 bp in length
2200   2299: gap of unknown length
2300   3670: contig of 1371 bp in length
3671   5275: gap of unknown length
5276   5375: contig of 1505 bp in length
5376   6409: gap of unknown length
6410   6509: contig of 1034 bp in length
6510   7565: gap of unknown length
7566   7665: contig of 1056 bp in length
7666   8816: gap of unknown length
8817   8916: contig of 1151 bp in length
8917   10099: gap of unknown length
10099  10199: contig of 1183 bp in length
10199  11446: gap of unknown length
11446  12748: contig of 1247 bp in length
12748  12848: gap of unknown length
12848  14094: contig of 1202 bp in length
14094  14194: gap of unknown length
14194  15241: contig of 1246 bp in length
15241  15341: gap of unknown length
15341  16738: contig of 1047 bp in length
16738  16838: gap of unknown length
16838  18428: contig of 1397 bp in length
18428  18528: gap of unknown length
18528  19576: contig of 1590 bp in length
19576  19676: gap of unknown length
19676  21464: contig of 1048 bp in length
21464  21564: gap of unknown length
21564  23696: contig of 2132 bp in length
23696  25356: gap of unknown length
25356  25656: contig of 1760 bp in length
25656  26780: gap of unknown length
26780  26880: contig of 1124 bp in length
26880  gap of unknown length

```

```

26881  28222: contig of 1342 bp in length
28222  28322: gap of unknown length
28322  29760: contig of 1438 bp in length
29760  29860: gap of unknown length
29860  31634: contig of 1774 bp in length
31634  31735: gap of unknown length
31735  33237: contig of 1502 bp in length
33237  35287: gap of unknown length
35287  35387: contig of 1951 bp in length
35387  38093: gap of unknown length
38093  38194: contig of 2706 bp in length
38194  40557: gap of unknown length
40557  40657: contig of 2364 bp in length
40657  44259: gap of unknown length
44259  44359: contig of 3602 bp in length
44359  46116: gap of unknown length
46116  46216: contig of 1757 bp in length
46216  48047: gap of unknown length
48047  48147: contig of 1831 bp in length
48147  50032: gap of unknown length
50032  50132: contig of 1885 bp in length
50132  52335: gap of unknown length
52335  52435: contig of 2203 bp in length
52435  54503: gap of unknown length
54503  54603: contig of 2067 bp in length
54603  57315: gap of unknown length
57315  57415: contig of 2713 bp in length
57415  59726: gap of unknown length
59726  59826: contig of 2311 bp in length
59826  63021: gap of unknown length
63021  63121: contig of 3195 bp in length
63121  65230: gap of unknown length
65230  65330: contig of 2109 bp in length
65330  67944: gap of unknown length
67944  70633: contig of 2614 bp in length
70633  70733: gap of unknown length
70733  73064: contig of 2330 bp in length
73064  73163: gap of unknown length
73163  76481: contig of 3318 bp in length
76481  78389: gap of unknown length
78389  78489: contig of 1808 bp in length
78489  82944: gap of unknown length
82944  83044: contig of 4454 bp in length
83044  86661: gap of unknown length
86661  86761: contig of 3618 bp in length
86761  89911: gap of unknown length
89911  90011: contig of 3150 bp in length
90011  94436: gap of unknown length
94436  94536: contig of 4425 bp in length
94536  98167: gap of unknown length
98167  98267: contig of 3631 bp in length
98267  102900: gap of unknown length
102900  103000: contig of 4633 bp in length
103000  107581: gap of unknown length
107581  107681: contig of 4581 bp in length
107681  110850: gap of unknown length
110850  110950: contig of 3169 bp in length
110950  113470: gap of unknown length
113470  113570: contig of 2520 bp in length
113570  117341: gap of unknown length
117341  117441: contig of 3771 bp in length
117441  121926: gap of unknown length
121926  122026: contig of 4485 bp in length
122026  124628: gap of unknown length
124628  124728: contig of 2602 bp in length
124728  132125: gap of unknown length
132125  contig of 7397 bp in length

```

Query Match 70.0%; Score 21; DB 2; Length 156128;
Best Local Similarity 82.8%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

HTG; HTGS_PHASE1; HTGS_DRAFT.

Homo sapiens

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 17876)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Albrock, S.L., Amaral, H.C., Are, J.R., Banks, T., Barbata, J., Benton, J., Bimaga, K., Blankenbury, K., Bonin, D., Bouck, J., Bowe, S., Brieve, M., Brown, E., Brown, M., Bryant, N.P., Buhy, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carion, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chokhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dethorne, S.R., David, R., Davila, M.L., Davis, C., Deen, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Drepper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hayes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Hollway, C., Hollins, B., Homsl, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kravtsov, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtenberg, O., Liu, C., Liu, J., Liu, W., Lousegod, H., Lozano, R.U., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newbott, K., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkwo, S., Ogulu, M., Okunnu, G., Otagunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Ruiters, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Russell, S., Saverly, G., Scher, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swatek, A., Tabor, P., Tameisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Tellford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished

2 (bases 1 to 17876)

Worley, K.C.

Submitted (07-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jun 25, 2001 this sequence version replaced gi:12831249.

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: HAMH

Center clone name: RP11-533P2

Summary Statistics

Sequencing vector: M13; L08821

Chemistry: Dye-terminator Big Dye; 99% of reads

Assembly program: Phrap; variation 0.990329

Consensus quality: 169802 bases at least Q40

Consensus quality: 180431 bases at least Q30

Consensus quality: 185136 bases at least Q20

Estimated insert size: 177736; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 4.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 42878: contig of 42878 bp in length
* 42879 42978: gap of unknown length
* 42979 42978: contig of 19040 bp in length
* 62019 62118: gap of unknown length
* 62119 80603: contig of 18485 bp in length
* 80604 80704: gap of unknown length
* 80704 98288: contig of 17585 bp in length
* 98289 98389: gap of unknown length
* 98389 113934: contig of 15546 bp in length
* 113935 114034: gap of unknown length
* 114035 125339: contig of 11305 bp in length
* 125340 125439: gap of unknown length
* 125440 138329: contig of 12890 bp in length
* 138330 138429: gap of unknown length
* 138430 145577: contig of 7148 bp in length
* 145578 145677: gap of unknown length
* 145678 150952: contig of 5275 bp in length
* 150953 151052: gap of unknown length
* 151053 155923: contig of 4871 bp in length
* 155924 156023: gap of unknown length
* 156024 159749: contig of 3725 bp in length
* 159750 159848: gap of unknown length
* 159849 163146: contig of 3298 bp in length
* 163147 163246: gap of unknown length
* 163247 16618: contig of 3372 bp in length
* 16619 166718: gap of unknown length
* 16672 170864: contig of 4146 bp in length
* 170865 170964: gap of unknown length
* 170965 173099: contig of 2135 bp in length
* 173100 173199: gap of unknown length
* 173200 175157: contig of 2318 bp in length
* 175158 175617: gap of unknown length
* 175618 17876: contig of 2259 bp in length.

FEATURES
source

1. 17876
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-533P2"
BASE COUNT 45021 a 43543 c 42200 g 45483 t 1629 others
ORIGIN

Query Match 70.0% Score 21; DB 2; Length 17876;
Best Local Similarity 82.8%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGACTCACTCTGTGTCATCTACCTCTCT 30
DB 68919 GGACTCACTCTGTGTCATCTACCTCTCT 68891

RESULT 14

AC130278

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

182695 bp DNA linear HTG 09-AUG-2002
unordered pieces.
AC130278
AC130278.1 GI:22165183
HTG; HTGS_PHASE1; HTGS_DRAFT.

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 182695)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP24-393116
Unpublished
2 (bases 1 to 182695)
Birren, B., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barra, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgeert, B.,
Canarata, J., Chang, J., Chararo, B., Choepell, Y., Collimore, A.,
Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-pierre, N., Hagos, B.,
Horton, L., Hulme, M., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Katsas, A., Kells, C., Landers, T., Levine, R., Lindblad-toh, K.,
Li, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Teafaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (09-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: 125607
Center clone name: 393.1.16

Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 168477 bases at least Q40
Consensus quality: 173980 bases at least Q30
Consensus quality: 176349 bases at least Q20
Insert size: 176000; agarose-ftp
Insert size: 179195; sum-of-coverage
Quality coverage: 5.6 in Q20 bases; agarose-ftp
Quality coverage: 5.6 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

I 7168: contig of 7168 bp in length
* 7169 7268: gap of 100 bp
* 7269 7773: contig of 505 bp in length
* 7774 7873: gap of 100 bp
* 7874 8549: contig of 676 bp in length
* 8550 8649: gap of 100 bp
* 8650 9264: contig of 615 bp in length
* 9265 9364: gap of 100 bp
* 9365 10058: contig of 694 bp in length
* 10059 10158: gap of 100 bp
* 10159 10902: contig of 744 bp in length
* 10903 11002: gap of 100 bp
* 11003 11256: contig of 254 bp in length
* 11257 11356: gap of 100 bp
* 11357 12233: contig of 877 bp in length

12234 12333: gap of 100 bp
* 12334 13682: contig of 1349 bp in length
* 13683 13782: gap of 100 bp
* 13783 14674: contig of 1092 bp in length
* 14675 14974: gap of 100 bp
* 14975 16211: contig of 1237 bp in length
* 16212 16311: gap of 100 bp
* 16312 17319: contig of 1008 bp in length
* 17320 17419: gap of 100 bp
* 17420 19489: contig of 2070 bp in length
* 19490 19589: gap of 100 bp
* 19590 21253: contig of 1664 bp in length
* 21254 21353: gap of 100 bp
* 21354 22801: contig of 1448 bp in length
* 22802 22901: gap of 100 bp
* 22902 24454: contig of 1533 bp in length
* 24455 24554: gap of 100 bp
* 24555 25711: contig of 1157 bp in length
* 25712 25811: gap of 100 bp
* 25812 28015: contig of 2204 bp in length
* 28016 28115: gap of 100 bp
* 28116 29791: contig of 1676 bp in length
* 29792 29891: gap of 100 bp
* 29892 32114: contig of 2223 bp in length
* 32115 32214: gap of 100 bp
* 32215 32974: contig of 760 bp in length
* 32975 33074: gap of 100 bp
* 33075 33793: contig of 719 bp in length
* 33794 33893: gap of 100 bp
* 33894 37124: contig of 3321 bp in length
* 37125 37224: gap of 100 bp
* 37225 40484: contig of 3260 bp in length
* 40485 40584: gap of 100 bp
* 40585 42991: contig of 2407 bp in length
* 42992 43091: gap of 100 bp
* 43092 48293: contig of 5202 bp in length
* 48294 48393: gap of 100 bp
* 48394 53122: contig of 4729 bp in length
* 53123 53222: gap of 100 bp
* 53223 75715: contig of 22493 bp in length
* 75716 75815: gap of 100 bp
* 75816 83666: contig of 7851 bp in length
* 83667 83766: gap of 100 bp
* 83767 92545: contig of 8779 bp in length
* 92546 92645: gap of 100 bp
* 92646 103351: contig of 10706 bp in length
* 103352 103451: gap of 100 bp
* 103452 118198: contig of 14747 bp in length
* 118199 118298: gap of 100 bp
* 118299 134605: contig of 16307 bp in length
* 134606 134705: gap of 100 bp
* 134706 151709: contig of 17004 bp in length
* 151710 151809: gap of 100 bp
* 151810 175524: contig of 23715 bp in length
* 175525 175624: gap of 100 bp
* 175625 182695: contig of 7071 bp in length.

Location/Qualifiers
1. 182695
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-393116"
/clone_lib="RP24 Male Mouse BAC"
1. 7168
/note="assembly-fragment"
vector_side:left"
clone_end:SP6
1. 7168
/note="assembly-fragment"
7269. 7773
/note="assembly-fragment"
7874. 8549
/note="assembly-fragment"
8650. 9264
/note="assembly-fragment"
9365. 10058

FEATURES
source
misc-feature
misc-feature
misc-feature
misc-feature
misc-feature
misc-feature
misc-feature

```

misc_feature /note="assembly-fragment"
10159..10902
/note="assembly-fragment"
misc_feature 11003..11256
/note="assembly-fragment"
misc_feature 11357..12233
/note="assembly-fragment"
misc_feature 12334..13682
/note="assembly-fragment"
misc_feature 13783..14874
/note="assembly-fragment"
misc_feature 14975..16211
/note="assembly-fragment"
misc_feature 16312..17319
/note="assembly-fragment"
misc_feature 17420..19489
/note="assembly-fragment"
misc_feature 19590..21253
/note="assembly-fragment"
misc_feature 21354..22801
/note="assembly-fragment"
misc_feature 22902..24454
/note="assembly-fragment"
misc_feature 24555..25711
/note="assembly-fragment"
misc_feature 25812..28015
/note="assembly-fragment"
misc_feature 28116..29791
/note="assembly-fragment"
misc_feature 29892..32114
/note="assembly-fragment"
misc_feature 32215..332974
/note="assembly-fragment"
misc_feature 33075..33793
/note="assembly-fragment"
misc_feature 33894..37124
/note="assembly-fragment"
misc_feature 37225..40484
/note="assembly-fragment"
misc_feature 40385..44291
/note="assembly-fragment"
misc_feature 43092..48293
/note="assembly-fragment"
misc_feature 48394..53122
/note="assembly-fragment"

```

```

Query Match 70.0%; Score 21; DB 2; Length 182695;
Best Local Similarity 82.8%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

Db 1 TGGACCTCAGTCCTGTCATCTCACCCTTC 29
51988 TGGCTCTCTGCTGTCATCTCACCCTTC 52016

RESULT 15
AC024545 193700 bp DNA linear HTG 23-SEP-2000
LOCUS AC024545/c Homo sapiens clone Rpl1-442P6, WORKING DRAFT SEQUENCE, 39 unordered
DEFINITION Pieces.
AC024545 AC024545.4 GI:10280888
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 193700)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193700)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

```

TITLE JOURNAL COMMENT

Anderson,S., Baldwin,J., Barna,N., Bede,F., Boguslavsky,L.,
Bouhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Klein,J., Lander,E., Lilev,I., Johnson,R., Jones,C., Kam,L., Karatas,A.,
Klein,J., Landers,T., Larcocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihov,T., Miranda,C., Mlenga,V., Morrow,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Plietner,N., Pisanl,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testave,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,C., Zainoun,J., Zimmer,A. and
Zody,M.

Direct Submission

Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 23, 2000 this sequence version replaced gi:8389572.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
Project Information

Center project name: L7456

Center clone name: L7456
Sequencing Statistics
Sequencing vector: MJ3; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 167380 bases at least Q40
Consensus quality: 180250 bases at least Q30
Consensus quality: 185744 bases at least Q20
Insert size: 188000; agarose-ff
Insert size: 189900; sum-of-ctnigs
Quality coverage: 3.3 in Q20 bases; agarose-ff
Quality coverage: 3.3 in Q20 bases; sum-of-ctnigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 1682: contig of 1682 bp in length
* 1683 1782: gap of 100 bp
* 1783 2982: contig of 1200 bp in length
* 2983 3082: gap of 100 bp
* 3083 4543: contig of 1461 bp in length
* 4544 4643: gap of 100 bp
* 4644 5643: contig of 1000 bp in length
* 5644 5743: gap of 100 bp
* 5744 7289: contig of 1546 bp in length
* 7290 7389: gap of 100 bp
* 7390 8911: contig of 1522 bp in length
* 8912 9011: gap of 100 bp
* 9012 11431: contig of 2420 bp in length
* 11432 11531: gap of 100 bp
* 11532 13523: contig of 1992 bp in length
* 13524 13623: gap of 100 bp
* 13624 16363: contig of 2740 bp in length
* 16364 16463: gap of 100 bp

```

```

* 16464 18732: contig of 2269 bp in length
* 18733 18832: gap of 100 bp
* 18833 21778: contig of 2946 bp in length
* 21779 21878: gap of 100 bp
* 21879 25118: contig of 3240 bp in length
* 25119 25218: gap of 100 bp
* 25219 27753: contig of 2335 bp in length
* 27754 27853: gap of 100 bp
* 27854 30849: contig of 2896 bp in length
* 30850 30949: gap of 100 bp
* 30950 33682: contig of 2733 bp in length
* 33683 33782: gap of 100 bp
* 33783 36641: contig of 2859 bp in length
* 36642 36741: gap of 100 bp
* 36742 40978: contig of 4237 bp in length
* 40979 41078: gap of 100 bp
* 41079 45552: contig of 4474 bp in length
* 45553 45652: gap of 100 bp
* 45653 48341: contig of 2689 bp in length
* 48342 48441: gap of 100 bp
* 48442 53528: contig of 5087 bp in length
* 53529 53628: gap of 100 bp
* 53629 59346: contig of 5718 bp in length
* 59347 59446: gap of 100 bp
* 59447 64005: contig of 4559 bp in length
* 64006 64105: gap of 100 bp
* 64106 68144: contig of 4039 bp in length
* 68145 68244: gap of 100 bp
* 68245 73407: contig of 5163 bp in length
* 73408 73507: gap of 100 bp
* 73508 77763: contig of 4256 bp in length
* 77764 77863: gap of 100 bp
* 77864 82554: contig of 4691 bp in length
* 82555 82654: gap of 100 bp
* 82655 95707: contig of 13053 bp in length
* 95708 95807: gap of 100 bp
* 95808 100939: contig of 5132 bp in length
* 100940 101039: gap of 100 bp
* 101040 106431: contig of 5392 bp in length
* 106432 106531: gap of 100 bp
* 106532 114394: contig of 7863 bp in length
* 114395 114494: gap of 100 bp
* 114495 120736: contig of 6242 bp in length
* 120737 120836: gap of 100 bp
* 120837 127881: contig of 7045 bp in length
* 127882 127981: gap of 100 bp
* 127982 136274: contig of 8293 bp in length
* 136275 136374: gap of 100 bp
* 136375 145186: contig of 8812 bp in length
* 145187 145286: gap of 100 bp
* 145287 153548: contig of 8262 bp in length
* 153549 153648: gap of 100 bp
* 153649 163735: contig of 10107 bp in length
* 163736 163855: gap of 100 bp
* 163856 172952: contig of 9097 bp in length
* 172953 173052: gap of 100 bp
* 173053 187116: contig of 14064 bp in length
* 187117 187216: gap of 100 bp
* 187217 193700: contig of 6484 bp in length.

```

FEATURES

Location/Qualifiers

```

1..193700
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-442P6"
/clone_lib="RPCT-11 Human Male BAC"
1..1682
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature
1783..2982
/note="assembly_fragment"
3083..4543
/note="assembly_fragment"

```

```

misc_feature 4644..5643
/note="assembly_fragment"
misc_feature 5744..7289
/note="assembly_fragment"
misc_feature 7390..8911
/note="assembly_fragment"
misc_feature 9012..11431
/note="assembly_fragment"
misc_feature 11532..13523
/note="assembly_fragment"
misc_feature 13624..16363
/note="assembly_fragment"
misc_feature 16464..18732
/note="assembly_fragment"
misc_feature 18833..21778
/note="assembly_fragment"
misc_feature 21879..25118
/note="assembly_fragment"
misc_feature 25219..27753
/note="assembly_fragment"
misc_feature 27854..30849
/note="assembly_fragment"
misc_feature 30950..33682
/note="assembly_fragment"
misc_feature 33783..36641
/note="assembly_fragment"
misc_feature 36742..40978
/note="assembly_fragment"
misc_feature 41079..45552
/note="assembly_fragment"
misc_feature 45653..48341
/note="assembly_fragment"
misc_feature 48442..53528
/note="assembly_fragment"
misc_feature 53629..59346
/note="assembly_fragment"
misc_feature 59447..64005
/note="assembly_fragment"

```

```

Query Match 70.0%; Score 21; DB 2; Length 193700;
Best Local Similarity 82.8%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 GGACGACGTCCTGTCATCCACCTCT 30
DB 67684 GGACGACGTCCTGTCATCCCTTCT 67656

```

Search completed: July 10, 2003, 19:20:11
Job time : 269.955 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 69.6337 Seconds
(without alignments)
1067.241 Million cell updates/sec

Title: US-10-087-631B-3

Perfect score: 33
Sequence: 1 cgggtaccacccgctccgagacacactatgyc 33

Scoring table: IDENTITY_NUC
Gapex 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	30.4	92.1	36	17	AAAT39735
C 2	27.4	83.0	29	19	AAAT15324
C 3	22	66.7	41	24	AAAL40116
C 4	22	66.7	68	20	AAAX60950
C 5	22	66.7	68	20	AAAX60951
C 6	22	66.7	70	17	AAAT11268
C 7	22	66.7	120	17	AAAT09180
C 8	22	66.7	120	18	AAAT69054
C 9	22	66.7	140	17	AAAT11269

ID	Score	Query Match	Length	ID	Description
10	66.7	145	17	AAAT09172	Hepatitis C virus
11	66.7	145	24	ABK6834	Hepatitis C virus
12	66.7	159	24	AA043062	-255 to -62 portio
13	66.7	159	14	AA043066	-255 to -62 region
14	66.7	159	14	AA043069	-255 to -62 region
15	66.7	159	14	AA043071	-255 to -62 region
16	66.7	163	22	AAAS10487	HCV 5'-UTR domain
17	66.7	177	15	AAAT09448	HCV isolate BE82 5
18	66.7	177	15	AAAT09449	HCV isolate BE90 5
19	66.7	177	15	AAAT09450	HCV isolate BE91 5
20	66.7	177	15	AAAT09451	HCV isolate BE92 5
21	66.7	177	15	AAAT09452	HCV isolate BE93 5
22	66.7	177	15	AAAT09453	HCV isolate BE94 5
23	66.7	177	15	AAAT09454	HCV isolate BE98 5
24	66.7	177	15	AAAT09455	HCV isolate GB48 5
25	66.7	177	15	AAAT09456	HCV isolate GB116
26	66.7	177	15	AAAT09457	HCV isolate GB569
27	66.7	177	15	AAAT09458	HCV isolate GB358
28	66.7	177	15	AAAT09459	HCV isolate CAM600
29	66.7	177	15	AAAT09460	HCV isolate CAM736
30	66.7	177	15	AA068063	HCV isolate GB809
31	66.7	177	15	AA068067	HCV isolate BE97 5
32	66.7	177	15	AA068068	HCV isolate GB487
33	66.7	177	15	AA068069	HCV isolate BE95 5
34	66.7	177	15	AA068070	HCV isolate BE96 5
35	66.7	180	13	AA031082	HCV-1 genotype GV
36	66.7	180	13	AA031083	HCV-1 genotype GV
37	66.7	184	14	AA043060	-255 to -62 portio
38	66.7	184	14	AA043061	-255 to -62 portio
39	66.7	184	14	AA043063	-255 to -62 region
40	66.7	184	14	AA043064	-255 to -62 region
41	66.7	184	14	AA043065	-255 to -62 region
42	66.7	184	14	AA043068	-255 to -62 region
43	66.7	187	14	AA043059	-255 to -62 portio
44	66.7	187	14	AA043065	-255 to -62 region
45	66.7	194	14	AA043058	-255 to -62 portio

ALIGNMENTS

RESULT 1

AAAT39735/c

ID AAAT39735 standard; DNA, 36 BP.

AC AAAT39735;

DT 09-APR-1997 (first entry)

DE Hepatitis C virus PCR probe S1, based on nucleotides 121-157.

XX Hepatitis C virus; HCV, polymerase chain reaction; amplification;

KW replication; non-lymphoblastoid cell; monkey kidney cell;

KM hybridisation probe; ss.

XX Synthetic.

OS

XX WO9624662-A1.

PN 15-AUG-1996.

PD 10-FEB-1995; 95WO-IT00016.

PE 10-FEB-1995; 95WO-IT00016.

PR 10-FEB-1995; 95WO-IT00016.

XX (CNDR) CONSIGLIO NAZ DELLE RICERCHE.

PA Battaglia M, Carloni G, Iacovacci S, Ponzetto A;

PI Ravagnan G;

XX WPI: 1996-384435/38.

DR Replication of hepatitis C virus in non-lymphoblastoid mammalian

XX PT

PT cells - useful for studies of HCV replication, prodn. of vaccines or
XX viral antigens, etc
XX
PS Disclosure; Page 8; 25pp; English.
XX
CC Hepatitis C virus can be replicated in non-lymphoblastoid mammalian
CC cells by first incubating an HCV sample with the cells until an
CC infecting amount of HCV has been absorbed. Infected cells are then
CC washed and incubated under growth conditions. In an example, the
CC presence of HCV in culture medium of monkey kidney cells inoculated
CC with HCV-infected serum was verified by PCR amplification using two
CC external primers (OU1 and OU2, see AAT39731 and AAT39732) and two
CC internal primers (IN1 and IN2, see AAT39731 and AAT39732). The amplified
CC products were identified by hybridisation to labelled probe S1 (see
CC AAT39733). Small amounts of virus were shown to be released into the
CC culture medium from secondary cultures. Control (non-inoculated) cells
CC were negative for presence of viral genome.
XX
SQ Sequence 36 BP; 9 A; 9 C; 11 G; 7 T; 0 other;
Query Match 92.1%; Score 30.4; DB 17; Length 36;
Best Local Similarity 96.9%; Pred. No. 0.00033;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 CGGTGTACTCACCCTCCGACGACCACTATGG 32
DB 32 CGGTGTACTCACCCTCCGACGACCACTATGG 1
RESULT 2
AAV15324/C
ID AAV15324 standard; DNA; 29 BP.
AC AAV15324;
XX
DT 28-MAY-1998 (first entry)
XX
DE Hepatitis C virus probe HCV40.
XX
KW Hepatitis C virus; HCV; PCR; detection; reverse transcription; probe;
XX enzyme immunoassay; viral RNA; ss.
XX
OS Synthetic.
XX Hepatitis C virus.
XX
PN WC9746716-A1.
XX
PD 11-DEC-1997
XX
PF 03-JUN-1997; 97WO-IT00128.
XX
PR 07-JUN-1996; 96IT-MO00404.
XX
PA (WESA) WABCO BV.
XX
PI Bosio P, Clemenza F, Strumia C;
XX WPI; 1998-042222/04.
XX
DR
XX
PT Detection of hepatitis C virus - by reverse transcription,
PT single-step PCR and detection by DNA enzyme immunoassay
XX
PS Disclosure; Page 4; 26pp; English.
XX
CC The present sequence represents a probe involved in the method of
CC the present invention for detecting hepatitis C virus (HCV). The method
CC comprises: (a) reverse-transcribing the viral RNA; (b) amplifying the
CC resulting cDNA by a single polymerase chain reaction in a reaction
CC mixture having a Mg2+/Taq polymerase ratio of about 100 nmole/enzyme
CC unit; and (c) detecting the amplification product by DEIA (DNA enzyme
CC immunoassay) using an oligonucleotide probe. The sensitivity of this
CC method is at least equal to that achievable by more complicated assays
CC using nested PCR.

XX
SQ Sequence 29 BP; 7 A; 7 C; 9 G; 6 T; 0 other;
Query Match 83.0%; Score 27.4; DB 19; Length 29;
Best Local Similarity 96.6%; Pred. No. 0.0072;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 4 TGTACTCACCCTCCGACGACCACTATGG 32
DB 29 TGTACTCACCCTCCGACGACCACTATGG 1
RESULT 3
AAL40116
ID AAL40116 standard; DNA; 41 BP.
AC AAL40116;
XX
DT 13-SEP-2002 (first entry)
XX
DE Pathogenic microorganism detecting PCR primer SEQ ID No 42.
XX
KW Tuberculosis; Mycobacterium; bovis; BCG; africanum; microti; canottii;
XX monitoring therapy; pathogenic microorganism; PCR; primer; ss.
XX
OS Unidentified.
XX
PN WO200252043-A1.
XX
PD 04-JUL-2002.
XX
PF 26-DEC-2001; 2001MO-JP11422.
XX
PR 26-DEC-2000; 2000JP-0396222.
XX
PR 26-DEC-2000; 2000JP-0396321.
XX
PR 29-JUN-2001; 2001JP-0199552.
XX
PR 13-SEP-2001; 2001JP-0278920.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Shinada M, Hino F, Kato I;
XX WPI; 2002-500769/53.
XX
DR
XX
PT Detecting pathogenic microorganisms with oligonucleotide probes and
PT primers, useful in disease diagnosis and monitoring therapy
XX
PS Claim 57; Page 97; 106pp; Japanese.
XX
CC The invention relates to a probe containing a 410 or 20 base pair
CC sequence, given in the specification. It is capable of detecting the
CC tuberculosis bacterial group including Mycobacterium tuberculosis,
CC Mycobacterium bovis BCG, Mycobacterium africanum, Mycobacterium microti
CC and/or Mycobacterium canottii. The method is useful in disease diagnosis
CC and monitoring therapy. This polynucleotide sequence represents a PCR
CC primer relating to the detection of pathogenic microorganisms of the
CC invention.
XX
SQ Sequence 41 BP; 7 A; 14 C; 9 G; 11 T; 0 other;
Query Match 66.7%; Score 22; DB 24; Length 41;
Best Local Similarity 97.1%; Pred. No. 2;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 1 CGGTGTACTCACC-GTTCGACGACCACTATGGC 33
DB 5 CGGTGTACTCACC-GTTCGACGACCACTATGGC 38
RESULT 4
AAK60950/C
ID AAK60950 standard; DNA; 68 BP.
XX

AC AAX60950;
 XX
 PD 16-AUG-1999 (first entry)
 XX
 DE Hepatitis C virus (HCV) biotinylated probe 3CH.
 XX
 DE Nucleic acid detection; pathogen; bacteria; virus; hepatitis C virus;
 XX HCV; hepatitis B; HBV; hepatitis G; HGV; HIV; fungus; protozoa; ss;
 XX parasite; mycoplasma; genetic mutation; food contamination; probe.
 XX
 OS Synthetic.
 OS Hepatitis C virus.
 XX
 PM WO9928503-A1.
 XX
 PD 10-JUN-1999.
 XX
 PF 16-NOV-1998; 98WO-US24494.
 XX
 PR 03-DEC-1997; 97IT-RM00749.
 XX
 PA (DIAS-) DIASORIN INT INC.
 XX
 PI Mantero G, Priml D;
 XX
 DR WPI, 1999-371139/31.
 XX
 PT Detection of single-stranded polynucleotide analytes
 XX
 PS Claim 6; Page 65; 73pp; English.
 XX
 CC The invention relates to a new method for detection of single-stranded
 CC (ss) polynucleotide analytes that comprises using ss polynucleotide
 CC probes which hybridize to the analyte and are bound to a solid support
 CC where double-stranded (ds) polynucleotides are detected. The method can
 CC be used for detecting ss-PN analytes for the detection of pathogens such
 CC as bacteria, viruses such as hepatitis C (HCV), hepatitis B (HBV),
 CC hepatitis G (HGV), or HIV, fungi, protozoa, parasites or mycoplasma. The
 CC method can be used to detect the presence of genetic mutations which have
 CC diagnostic or prognostic value. Pathogen contamination of food and drink
 CC supplies can also be detected using the method. The method provides for
 CC the sensitive and specific detection of ss-PN analytes at concentration
 CC as low as 0.1fg.
 CC
 SQ Sequence 68 BP; 14 A; 17 C; 24 G; 13 T; 0 other;
 XX
 XX
 Query Match 66.7%; Score 22; DB 20; Length 68;
 Best Local Similarity 97.1%; Pred. No. 2.2;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 OY 1 CGGTGTACTCACCGTCCGACGACCACTATGCG 33
 DB 40 CGGTGTACTCACCGTCCGACGACCACTATGCG 7
 XX
 RESULT 5
 AAX60951
 ID AAX60951 standard; DNA; 68 BP.
 XX
 AC AAX60951;
 XX
 DT 16-AUG-1999 (first entry)
 XX
 DE Hepatitis C virus (HCV) biotinylated probe alpha-3CH.
 XX
 DE Nucleic acid detection; pathogen; bacteria; virus; hepatitis C virus;
 XX HCV; hepatitis B; HBV; hepatitis G; HGV; HIV; fungus; protozoa; ss;
 XX parasite; mycoplasma; genetic mutation; food contamination; probe.
 XX
 OS Synthetic.
 OS Hepatitis C virus.
 XX
 PM WO9928503-A1.

XX
 PD 10-JUN-1999.
 XX
 PF 16-NOV-1998; 98WO-US24494.
 XX
 PR 03-DEC-1997; 97IT-RM00749.
 XX
 PA (DIAS-) DIASORIN INT INC.
 XX
 PI Mantero G, Priml D;
 XX
 DR WPI, 1999-371139/31.
 XX
 PT Detection of single-stranded polynucleotide analytes
 XX
 PS Claim 6; Page 66; 73pp; English.
 XX
 CC The invention relates to a new method for detection of single-stranded
 CC (ss) polynucleotide analytes that comprises using ss polynucleotide
 CC probes which hybridize to the analyte and are bound to a solid support
 CC where double-stranded (ds) polynucleotides are detected. The method can
 CC be used for detecting ss-PN analytes for the detection of pathogens such
 CC as bacteria, viruses such as hepatitis C (HCV), hepatitis B (HBV),
 CC hepatitis G (HGV), or HIV, fungi, protozoa, parasites or mycoplasma. The
 CC method can be used to detect the presence of genetic mutations which have
 CC diagnostic or prognostic value. Pathogen contamination of food and drink
 CC supplies can also be detected using the method. The method provides for
 CC the sensitive and specific detection of ss-PN analytes at concentration
 CC as low as 0.1fg.
 CC
 SQ Sequence 68 BP; 13 A; 24 C; 17 G; 14 T; 0 other;
 XX
 XX
 Query Match 66.7%; Score 22; DB 20; Length 68;
 Best Local Similarity 97.1%; Pred. No. 2.2;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 OY 1 CGGTGTACTCACCGTCCGACGACCACTATGCG 33
 DB 29 CGGTGTACTCACCGTCCGACGACCACTATGCG 62
 XX
 RESULT 6
 AAT11268
 ID AAT11268 standard; RNA; 70 BP.
 XX
 AC AAT11268;
 XX
 DT 26-JUN-1996 (first entry)
 XX
 DE Hepatitis C virus partial 5'-UTR antisense RNA AS3.
 XX
 DE Antisense; therapy; complementary; HCV; 5'-untranslated region;
 XX hepatitis C virus; inhibition; infection; treatment; stem-loop;
 XX clone 2-1; ss.
 XX
 OS Hepatitis C virus.
 OS
 PM JP07303485-A.
 XX
 PD 21-NOV-1995.
 XX
 PF 13-MAY-1994; 94UP-0124609.
 XX
 PR 13-MAY-1994; 94JP-0124609.
 XX
 PA (TOFU) TONEN CORP.
 XX
 DR WPI, 1996-035187/04.
 XX
 PT Hepatitis C virus (HCV) anti-sense RNA - inhibits HCV structural
 PT gene expression in vivo for treatment of HCV infection
 XX
 PS Claim 2; Page 9; 12pp; Japanese.

XX The present sequence is a specifically claimed example of RNA that
 CC is complementary (i.e. antisense) to part of the 5' untranslated
 CC region of the hepatitis C virus genome sequence contained in clone
 CC 2-1. The 5'-UTR includes several stem-loop sequences. The antisense
 CC RNA is useful for inhibiting expression of HCV structural genes and
 CC thereby inhibiting viral replication in vivo. The antisense therapy
 CC can be used in addition to conventional interferon treatment of HCV
 CC infections.

XX Sequence 70 BP; 10 A; 21 C; 25 G; 14 U; 0 other;

XX Query Match Best Local Similarity 66.7%; Score 22; DB 17; Length 70;
 XX Matches 26; Conservative 7; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGCGACGACACTATGCG 33
 Db 10 CGGUGNACUCACCGGUCGCGACACCAUAUGGC 43

RESULT 7

ID AAT09180 standard; DNA; 120 BP.

XX AAT09180;

XX 14-AUG-1996 (first entry).

XX Hepatitis C virus specific ligated amplified sequence.

XX ligase dependent polymerase chain reaction; LD-PCR; probe; hybridisation;
 KM ligand binding pair; ligase; paramagnetic bead; primer; amplification;
 KM hepatitis; untranslated region; UTR; RNA; ss.

XX Synthetic.

XX WO9353390-A1.

XX 28-DEC-1995.

XX 14-JUN-1995; 95WO-US07671.

XX 22-JUN-1994; 94US-0263937.

XX (MOUN) MOUNT SINAI SCHOOL MEDICINE.

XX Zhang DY;

XX WPI; 1996-058427/06.

XX Ligation dependent polymerase chain reaction - for the detection of
 PT infections pathogens and abnormal human genes, e.g. HIV and
 PT neoplasia

XX Example 5; Page 54; 100pp; English.

XX A novel method of detecting a target nucleic acid (TNA) sequence
 CC involves use of the ligase dependent polymerase chain reaction method
 CC (LD-PCR). In this method, two probes are provided. The first probe
 CC contains a region at the 5' end which is complementary and will hybridise
 CC with the TNA, the 3' end of the first probe is generic and is bound to
 CC one half of a ligand binding pair (LBP). The second probe contains a
 CC region at the 5' end which is complementary to a region in the TNA which
 CC is immediately adjacent to the complementary region of the first probe.
 CC When the probes are bound to the TNA, they can be ligated together using
 CC a conventional ligase. The TNA:ligated probe complex can be isolated by
 CC binding the first probe to a paramagnetic bead to which is attached the
 CC second half of the LBP. The TNA can be dissociated from the ligated
 CC probe complex which can then be detected either by a label attached to
 CC the second probe, by using an external probe or by PCR using the ligated
 CC probes as a template.
 CC The capture probes AAT09176-7 are used to isolated a region of the

CC Hepatitis C virus 5' untranslated region. This is the sequence of
 CC the ligated amplified region created by the ligation of the two
 CC amplification probes amp-probe-2 and -2A (AAT09178-9). The ligated
 CC sequence can subsequently be detected by PCR amplification with
 CC the primers AAT09181-3.

XX Sequence 120 BP; 26 A; 34 C; 29 G; 31 T; 0 other;

XX Query Match Best Local Similarity 66.7%; Score 22; DB 17; Length 120;
 XX Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGCGACGACACTATGCG 33
 Db 64 CGGTGTACTCACC-GTTCGCGACGACACTATGCG 97

RESULT 8

ID AAT69054/c standard; DNA; 120 BP.

XX AAT69054;

XX 16-JUL-1997 (first entry)

XX Hepatitis C virus fragment.

XX Polymerase chain reaction; amplification; hepatitis B virus; HBV;
 KM hepatitis C virus; HCV; simultaneous analysis; nested PCR; ss.

XX Synthetic.

XX KR9508291-B1.

XX 27-JUL-1995.

XX 22-APR-1992; 92KR-0006826.

XX 22-APR-1992; 92KR-0006826.

XX (MOUN/) MUN I.

XX Mun I;

XX WPI; 1997-191071/17.

XX Simultaneously analysing hepatitis B and C virus - using polymerase
 PT chain reaction

XX Example 3; Page 7; 12pp; Korean.

XX Hepatitis B virus (HBV) and hepatitis C virus (HCV) can be analysed
 CC simultaneously using specific sets of PCR primers. Specifically,
 CC viral nucleic acids are isolated by treating human serum with a
 CC guanidine isothiocyanate containing solution. HBV and HCV nucleic
 CC acids are then absorbed on silica and a first polymerase chain reaction
 CC is performed using primers HB-1 and HB-2 (AAT65089 and AAT65090) and
 CC HC-1 and HC-2 (AAT65093 and AAT65094). A second PCR is then performed
 CC with the primers HB-3 and HB-4 (AAT65091 and AAT65092) and HC-3 and HC-4
 CC (AAT65095 and AAT65096). A PCR fragment of 258 bp is produced for HBV and
 CC a PCR fragment of 163 bp is produced for HCV. The present sequence
 CC represents a fragment from HCV.

XX Sequence 120 BP; 21 A; 37 C; 34 G; 22 T; 6 other;

XX Query Match Best Local Similarity 66.7%; Score 22; DB 18; Length 120;
 XX Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGCGACGACACTATGCG 33
 Db 58 CGGTGTACTCACC-GTTCGCGACGACACTATGCG 25


```

RESULT 9
AAT11269
ID AAT11269 standard; RNA: 140 BP.
XX
AC AAT11269;
XX
DT 26-JUN-1996 (first entry)
XX
DE Hepatitis C virus partial 5'-UTR antisense RNA AS3'.
XX
KW Antisense; therapy; complementary; HCV; 5'-untranslated region;
KM hepatitis C virus; inhibition; infection; treatment; stem-loop;
KW clone 2-1; ss.
XX
OS Hepatitis C virus.
XX
PN JP07303485-A.
XX
PD 21-NOV-1995.
XX
PF 13-MAY-1994; 94JP-0124609.
XX
PR 13-MAY-1994; 94JP-0124609.
XX
PA (TOFU ) TONEN CORP.
XX
DR WPI; 1996-035187/04.
XX
PT Hepatitis C virus (HCV) antisense RNA - inhibits HCV structural
PT gene expression in vivo for treatment of HCV infection
XX
PS Claim 2; Page 9; 12pp; Japanese.
XX
CC The present sequence is a specifically claimed example of RNA that
CC is complementary (i.e. antisense) to part of the 5'-untranslated
CC region of the hepatitis C virus genome sequence contained in clone
CC 2-1. The 5'-UTR includes several stem-loop sequences. The antisense
CC RNA is useful for inhibiting expression of HCV structural genes and
CC thereby inhibiting viral replication in vivo. The antisense therapy
CC can be used in addition to conventional interferon treatment of HCV
CC infections.
XX
SQ Sequence 140 BP; 26 A; 43 C; 46 G; 25 U; 0 other;
XX
Query Match 66.7%; Score 22; DB 17; Length 140;
Best Local Similarity 76.5%; Pred. No. 2.4;
Matches 26; Conservative 7; Mismatches 0; Indels 1; Gaps 1;
OY 1 CGGTGTTACTCACC-GTTCGCGACGACACTATGCG 33
| | | | | | | | | | | | | | | | | | | | |
DB 80 CGGUGUACUCACCGGUGUCCGACAGCACUAGGCG 113

RESULT 10
AAT09172
ID AAT09172 standard; DNA: 145 BP.
XX
AC AAT09172;
XX
DT 14-AUG-1996 (first entry)
XX
DE Hepatitis C virus specific ligated amplification sequence.
XX
KW Ligase dependent polymerase chain reaction; LD-PCR; probe; hybridisation;
KM ligand binding pair; ligase; paramagnetic bead; primer; amplification;
KW hepatitis; untranslated region; UTR; RNA; ss.
XX
OS Synthetic.
XX
PN W09535390-A1.
XX
PD 28-DEC-1995.

```

```

XX
PF 14-JUN-1995; 95MO-US07671.
XX
PR 22-JUN-1994; 94US-0263937.
XX
PA (MOUN ) MOUNT SINAI SCHOOL MEDICINE.
XX
PI Zhang DY;
XX
DR WPI; 1996-058427/06.
XX
PT Ligation dependent polymerase chain reaction - for the detection of
PT infectious pathogens and abnormal human genes, e.g. HIV and
PT neoplasia
XX
PS Example 4; Page 51; 100pp; English.
XX
CC A novel method of detecting a target nucleic acid (TNA) sequence
CC involves use of the ligase dependent polymerase chain reaction method
CC (LD-PCR). In this method, two probes are provided. The first probe
CC contains a region at the 5' end which is complementary and will hybridise
CC with the TNA, the 3' end of the first probe is generic and is bound to
CC one half of a ligand binding pair (LBP). The second probe contains a
CC region at the 5' end which is complementary to a region in the TNA which
CC is immediately adjacent to the complementary region of the first probe.
CC When the probes are bound to the TNA, they can be ligated together using
CC a conventional ligase. The TNA:ligated probe complex can be isolated by
CC binding the first probe to a paramagnetic bead to which is attached the
CC second half of the LBP. The TNA can be dissociated from the ligated
CC probe complex which can then be detected either by a label attached to
CC the second probe, by using an external probe or by PCR using the ligated
CC probes as a template.
CC The probes AAT09170-1 are used to detect the presence of Hepatitis C
CC virus in a sample. This sequence results from the ligation of these
CC probes and contains sequence which is complementary to a region of the
CC hepatitis C virus 5' untranslated region. The sequence can be detected
CC by PCR using the primers AAT09159-60.
XX
SQ Sequence 145 BP; 29 A; 40 C; 40 G; 36 T; 0 other;
XX
Query Match 66.7%; Score 22; DB 17; Length 145;
Best Local Similarity 97.1%; Pred. No. 2.4;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 1 CGGTGTTACTCACC-GTTCGCGACGACACTATGCG 33
| | | | | | | | | | | | | | | | | | | | |
DB 72 CGGTGTTACTCACC-GTTCGCGACGACACTATGCG 105

RESULT 11
ABR86834
ID ABR86834 standard; DNA: 145 BP.
XX
AC ABR86834;
XX
DT 24-SEP-2002 (first entry)
XX
DE Hepatitis C virus ligated amplification sequence.
XX
KW RNA detection; ss; nucleic acid detection; infectious disease;
KW hepatitis.
XX
OS Hepatitis C virus.
XX
PN W0200244339-A2.
XX
PD 06-JUN-2002.
XX
PF 03-DEC-2001; 2001MO-US45822.
XX
PR 01-DEC-2000; 2000US-0728265.
XX
PA (ZHAN/) ZHANG D Y.

```

PA (BRAN/), BRANDWEIN M.
 PA (HSUI/), HSUI T C H.
 PI Zhang DY, Brandwein M, Hsu TH;
 DR WPI; 2002-508508/54.
 PT Detecting a target nucleic acid by a ramification-extension
 PT amplification method, involves carrying out the steps of capturing,
 PT carrying out ligand-dependent amplification and detection of the
 PT target nucleic acid
 PS Example 4; Page 78; 160pp; English.

CC This invention relates to a novel method for detecting target nucleic
 CC acid using a Capture/Amplification (C/A) probe and an amplification
 CC probe (A-probe), where C/A probe contains generic sequences and
 CC sequences complementary to target nucleic acid sequences and
 CC contains a ligand bound to the non-complementary sequence of probe; the
 CC ligand is capable of forming an affinity pair with a ligand binding
 CC moiety coated on matrix. The method of the invention may be used for
 CC detecting a target nucleic acid in sample. Preferably, the method is
 CC useful for detecting and quantitating nucleic acids from pathogenic
 CC microorganisms, from samples from patients with infectious diseases.
 CC The method allows for rapid, sensitive and standardised detection and
 CC quantification of nucleic acids from pathogenic microorganisms from
 CC the target nucleic acid isolation procedure, which can be performed in
 CC microtubes, microchips or micro-well plates, if desired. The method
 CC also allows the direct detection of RNA by probe amplification without
 CC the need for DNA template production. The assembly of the amplifiable
 CC DNA by ligation increases specificity, and makes possible the detection
 CC of a single mutation in a target. The method can be easily automated.
 CC The present sequence represents the nucleic acid detection method
 CC hepatitis C virus ligated amplification sequence used in an example of
 CC the method of the invention
 SO Sequence 145 BP; 29 A; 40 C; 40 G; 36 T; 0 other;

Query Match: 66.7%; Score 22; DB 24; Length 145;
 Best Local Similarity 97.1%; Pred. No. 2.4;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGACTACAC-GTTCGCGAGACCACTATGCG 33
 DB 72 CGGTGACTACACGCGTTCGCGAGACCACTATGCG 105

RESULT 12
 AAQ43062/C
 ID AAQ43062 standard; cDNA; 159 BP.

AC AAQ43062;
 DT 23-SEP-1993 (first entry)
 DE -255 to -62 portion of 5' non-coding region of HCV K2a.
 KW Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
 KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3; ss.
 OS Hepatitis C virus.
 PN WO9310239-A.
 PD 27-MAY-1993.
 PF 20-NOV-1992; 92MO-GB02143.
 PR 21-NOV-1991; 91GB-0024696.
 PR 24-JUN-1992; 92GB-0013362.
 (COMM-) COMMON SERVICES AGENCY.

XX Chan S, Simmonds P, Yap PL;
 PI WPI; 1993-182554/22.
 DR
 XX

PT DNA encoding antigenic peptide(s) of new types of hepatitis C
 PT virus - for diagnosing and treating HCV infection, screening
 PT blood samples and identifying different HCV types
 PS Disclosure; Fig 1; 120pp; English.

CC The sequences given in AAQ43058-75 show the -255 to -62 non-coding
 CC region of hepatitis C virus (HCV) samples from 18 blood donors and
 CC other HCV variants. Analysis of this region revealed the existence
 CC of three distinct groups of HCV differing by 9-14% in nucleotide
 CC sequence. Two of the groups identified were similar to those of HCV
 CC variants termed type 1 and 2, whilst the third appeared to represent
 CC a novel virus type. Comparison of other regions of the genome, eg.
 CC the NS-5 region (see also AAR37923-26), showed a high degree of sequence
 CC diversity with type 3 being phylogenetically different to type 1 and 2.
 CC The same degree of differentiation was noted in the NS-3 (see AAR37927-30)
 CC and core region (see AAR37931) between type 3 and type 1 sequences.

SO Sequence 159 BP; 29 A; 49 C; 48 G; 33 T; 0 other;
 Query Match: 66.7%; Score 22; DB 14; Length 159;
 Best Local Similarity 97.1%; Pred. No. 2.5;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGACTACAC-GTTCGCGAGACCACTATGCG 33
 DB 84 CGGTGACTACACGCGTTCGCGAGACCACTATGCG 51

RESULT 13
 AAQ43066/C
 ID AAQ43066 standard; cDNA; 159 BP.

AC AAQ43066;
 DT 23-SEP-1993 (first entry)
 DE -255 to -62 region of 5' non-coding region of HCV K2b-1.
 KW Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
 KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3; ss.
 OS Hepatitis C virus.
 PN WO9310239-A.
 PD 27-MAY-1993.
 PF 20-NOV-1992; 92MO-GB02143.
 PR 21-NOV-1991; 91GB-0024696.
 PR 24-JUN-1992; 92GB-0013362.
 (COMM-) COMMON SERVICES AGENCY.
 PI Chan S, Simmonds P, Yap PL;
 PI WPI; 1993-182554/22.
 DR
 XX DNA encoding antigenic peptide(s) of new types of hepatitis C
 PT virus - for diagnosing and treating HCV infection, screening
 PT blood samples and identifying different HCV types
 PS Disclosure; Fig 1; 120pp; English.
 CC The sequences given in AAQ43058-75 show the -255 to -62 non-coding
 CC region of hepatitis C virus (HCV) samples from 18 blood donors and
 CC other HCV variants. Analysis of this region revealed the existence

CC of three distinct groups of HCV differing by 9-14% in nucleotide
CC sequence. Two of the groups identified were similar to those of HCV
CC variants termed type 1 and 2, whilst the third appeared to represent
CC a novel virus type. Comparison of other regions of the genome, eg.
CC the NS-5 region (see also AAR31923-26), showed a high degree of sequence
CC diversity with type 3 being phylogenetically different to type 1 and 2.
CC The same degree of differentiation was noted in the NS-3 (see AAR31927-30)
CC and core region (see AAR31931) between type 3 and type 1 sequences.
XX
S0 Sequence 159 BP: 30 A: 48 C: 47 G: 34 T: 0 other:

Query Match	66.7%	Score 22:	DB 14;	Length 159;
Best Local Similarity	97.1%	Pred. No. 2.5;		
Matches 33;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1;

[illegible]

RESULT 14
AAQ43069/c
ID AAQ43069 standard; cDNA; 159 BP
vv

AC	AAQ43069;
XX	
DT	23-SEP-1993 (first entry)
YY	

-255 to -62 region of 5' non-coding region of HCV Pt-1.

KM Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
 KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3; ss
 vv
 ww

05 Hepatitis C virus.

PN W09310239-A.

PD 27-MAY-1993

PF 20-NOV-1992; 92WO-GB02143.
VY

PR	21-NOV-1991;	91GB-0024696.
PR	24-JUN-1992;	92GB-0013363

XX
XX
BA (COMM-) COMMON SERVICES AGENCY

PI Chan S, Simmonds P, Yap PL;

DR WPI; 1993-182554/.22.

PT DNA encoding antigenic peptide(s) of new types of hepatitis C
PT virus - for diagnosing and treating HCV infection, screening
PT blood samples and identifying different HCV types

PS Disclosure; Fig 1; 120pp; English

CC The sequences given in AA043055-75 show the -255 to -62 non-coding
CC region of hepatitis C virus (HCV) samples from 18 blood donors and
CC other HCV variants. Analysis of this region revealed the existence
CC of three distinct groups of HCV differing by 9-14% in nucleotide
CC sequence. Two of the groups identified were similar to those of HCV
CC variants termed type 1 and 2, whilst the third appeared to represent
CC a novel virus type. Comparison of other regions of the genome, eg.
CC the NS-5 region (see also AAR37923-26), showed a high degree of sequence
CC diversity with type 3 being phylogenetically different to type 1 and 2.
CC The same degree of differentiation was noted in the NS-3 (see AAR37927-30)
CC and core region (see AAR37931) between type 3 and type 1 sequences.

50 Sequence 159 BP; 30 A; 48 C; 50 G; 31 T; 0 other,

```

Query Match      66.7%;   Score 22;   DB 14;   length 159;
Best Local Similarity 97.1%;   Pred. No. 2.5;
Matches 33;   Conservative 0;   Indels 1;   Gaps 1.

```

27 1 CGGTGTTACTACC-GTTCGGCAGACCACTATGGC 33
 38 |||||
 49 |||||
 60 |||||
 71 |||||
 82 |||||
 93 |||||
 104 |||||
 115 |||||
 126 |||||
 137 |||||
 148 |||||
 159 |||||
 170 |||||
 181 |||||
 192 |||||
 203 |||||
 214 |||||
 225 |||||
 236 |||||
 247 |||||
 258 |||||
 269 |||||
 280 |||||
 291 |||||
 302 |||||
 313 |||||
 324 |||||
 335 |||||
 346 |||||
 357 |||||
 368 |||||
 379 |||||
 390 |||||
 401 |||||
 412 |||||
 423 |||||
 434 |||||
 445 |||||
 456 |||||
 467 |||||
 478 |||||
 489 |||||
 500 |||||
 511 |||||
 522 |||||
 533 |||||
 544 |||||
 555 |||||
 566 |||||
 577 |||||
 588 |||||
 599 |||||
 610 |||||
 621 |||||
 632 |||||
 643 |||||
 654 |||||
 665 |||||
 676 |||||
 687 |||||
 698 |||||
 709 |||||
 720 |||||
 731 |||||
 742 |||||
 753 |||||
 764 |||||
 775 |||||
 786 |||||
 797 |||||
 808 |||||
 819 |||||
 830 |||||
 841 |||||
 852 |||||
 863 |||||
 874 |||||
 885 |||||
 896 |||||
 907 |||||
 918 |||||
 929 |||||
 940 |||||
 951 |||||
 962 |||||
 973 |||||
 984 |||||
 995 |||||
 1006 |||||
 1017 |||||
 1028 |||||
 1039 |||||
 1050 |||||
 1061 |||||
 1072 |||||
 1083 |||||
 1094 |||||
 1105 |||||
 1116 |||||
 1127 |||||
 1138 |||||
 1149 |||||
 1160 |||||
 1171 |||||
 1182 |||||
 1193 |||||
 1204 |||||
 1215 |||||
 1226 |||||
 1237 |||||
 1248 |||||
 1259 |||||
 1270 |||||
 1281 |||||
 1292 |||||
 1303 |||||
 1314 |||||
 1325 |||||
 1336 |||||
 1347 |||||
 1358 |||||
 1369 |||||
 1380 |||||
 1391 |||||
 1402 |||||
 1413 |||||
 1424 |||||
 1435 |||||
 1446 |||||
 1457 |||||
 1468 |||||
 1479 |||||
 1490 |||||
 1501 |||||
 1512 |||||
 1523 |||||
 1534 |||||
 1545 |||||
 1556 |||||
 1567 |||||
 1578 |||||
 1589 |||||
 1600 |||||
 1611 |||||
 1622 |||||
 1633 |||||
 1644 |||||
 1655 |||||
 1666 |||||
 1677 |||||
 1688 |||||
 1699 |||||
 1710 |||||
 1721 |||||
 1732 |||||
 1743 |||||
 1754 |||||
 1765 |||||
 1776 |||||
 1787 |||||
 1798 |||||
 1809 |||||
 1820 |||||
 1831 |||||
 1842 |||||
 1853 |||||
 1864 |||||
 1875 |||||
 1886 |||||
 1897 |||||
 1908 |||||
 1919 |||||
 1930 |||||
 1941 |||||
 1952 |||||
 1963 |||||
 1974 |||||
 1985 |||||
 1996 |||||
 2007 |||||
 2018 |||||
 2029 |||||
 2040 |||||
 2051 |||||
 2062 |||||
 2073 |||||
 2084 |||||
 2095 |||||
 2106 |||||
 2117 |||||
 2128 |||||
 2139 |||||
 2150 |||||
 2161 |||||
 2172 |||||
 2183 |||||
 2194 |||||
 2205 |||||
 2216 |||||
 2227 |||||
 2238 |||||
 2249 |||||
 2260 |||||
 2271 |||||
 2282 |||||
 2293 |||||
 2304 |||||
 2315 |||||
 2326 |||||
 2337 |||||
 2348 |||||
 2359 |||||
 2370 |||||
 2381 |||||
 2392 |||||
 2403 |||||
 2414 |||||
 2425 |||||
 2436 |||||
 2447 |||||
 2458 |||||
 2469 |||||
 2480 |||||
 2491 |||||
 2502 |||||
 2513 |||||
 2524 |||||
 2535 |||||
 2546 |||||
 2557 |||||
 2568 |||||
 2579 |||||
 2590 |||||
 2601 |||||
 2612 |||||
 2623 |||||
 2634 |||||
 2645 |||||
 2656 |||||
 2667 |||||
 2678 |||||
 2689 |||||
 2700 |||||
 2711 |||||
 2722 |||||
 2733 |||||
 2744 |||||
 2755 |||||
 2766 |||||
 2777 |||||
 2788 |||||
 2799 |||||
 2810 |||||
 2821 |||||
 2832 |||||
 2843 |||||
 2854 |||||
 2865 |||||
 2876 |||||
 2887 |||||
 2898 |||||
 2909 |||||
 2920 |||||
 2931 |||||
 2942 |||||
 2953 |||||
 2964 |||||
 2975 |||||
 2986 |||||
 2997 |||||
 3008 |||||
 3019 |||||
 3030 |||||
 3041 |||||
 3052 |||||
 3063 |||||
 3074 |||||
 3085 |||||
 3096 |||||
 3107 |||||
 3118 |||||
 3129 |||||
 3140 |||||
 3151 |||||
 3162 |||||
 3173 |||||
 3184 |||||
 3195 |||||
 3206 |||||
 3217 |||||
 3228 |||||
 3239 |||||
 3250 |||||
 3261 |||||
 3272 |||||
 3283 |||||
 3294 |||||
 3305 |||||
 3316 |||||
 3327 |||||
 3338 |||||
 3349 |||||
 3360 |||||
 3371 |||||
 3382 |||||
 3393 |||||
 3404 |||||
 3415 |||||
 3426 |||||
 3437 |||||
 3448 |||||
 3459 |||||
 3470 |||||
 3481 |||||
 3492 |||||
 3503 |||||
 3514 |||||
 3525 |||||
 3536 |||||
 3547 |||||<

RESULT 15
AAQ43071/c
ID AAQ43071 standard; cDNA; 159 BP

AC AAQ43071;
yy

DT 23-SEP-1993 (first entry)
XY

DE -255 to -62 region of 5' non-coding region of HCV J1
XY

KM Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
KM HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3; ss

OS Hepatitis C virus

PN WO9310239-A.

PD 27-MAY-1993
YY

PF 20-NOV-1992; 92WO-GB02143
YY

PR	21-NOV-1991;	91GB-0024696
PR	24-JUN-1992;	92GB-0013362

XX
XX
(COMM-) COMMON SERVICES AGENCY

PI Chan S, Simmonds P, Yap PL;

DR WPI; 1993-182554/22.

PT DNA encoding antigenic peptide(s) of new types of hepatitis C
PT virus - for diagnosing and treating HCV infection, screening
PT blood samples and identifying different HCV types

PS Disclosure; Fig 1; 120pp; English

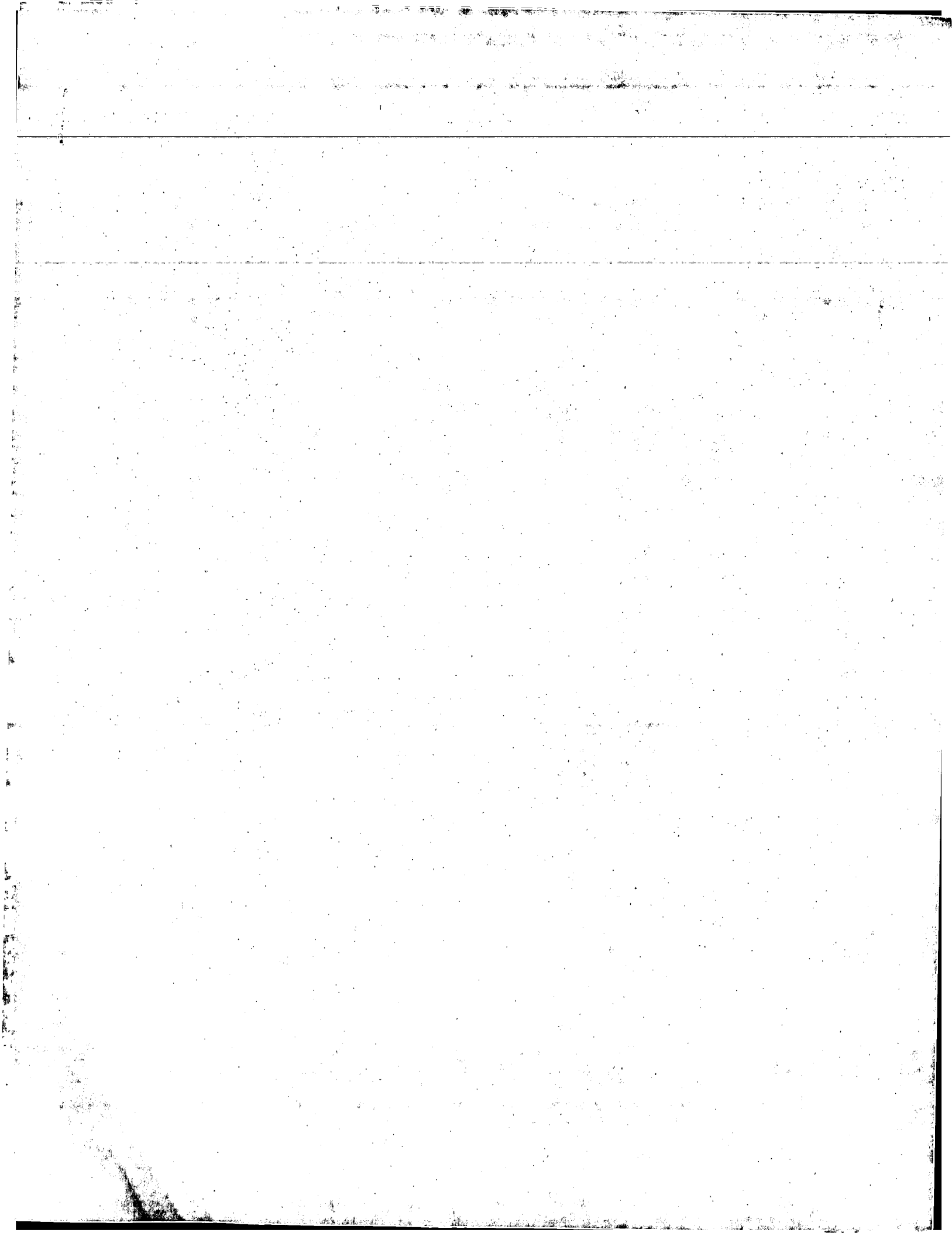
The sequences given in AAO43058-875 show the -255 to -62 non-coding region of hepatitis C virus (HCV) samples from 18 blood donors and other HCV variants. Analysis of this region revealed the existence of three distinct groups of HCV differing by 9-14% in nucleotide sequence. Two of the groups identified were similar to those of HCV variants termed type 1 and 2, whilst the third appeared to represent a novel virus type. Comparison of other regions of the genome, eg. the NS-5 region (see also AAR37923-26), showed a high degree of sequence diversity with type 3 being phylogenetically different to type 1 and 2. The same degree of differentiation was noticed in the NS-3 (see AAR37927-30) and core region (see AAR37931) between type 3 and type 1 sequences.

50 Sequence 159 BP; 30 A; 49 C; 50 G; 30 T; 0 other

Query Match	66.7%	Score 22:	DB 14;	Length 159;
Best Local Similarity	97.1%	Pred. No.	2.5;	
Matches 33;	Conservative	0;	Mismatches	1;
			Indels	1;
			Gaps	1;

QY 1 CGGTGTACTACACC-GATTCGGCAGACACCATATATGGC 3
 |||||
 DB 84 CGGTGTACTACACCGGATTCGGCAGACACCATATATGGC 57

Search completed: July 10, 2003, 19:52:33
Job time : 70.6337 secs



Typing nucleic acid for obtaining information about several variable

PT sites involves simultaneously or sequentially performing two or more
PT primer extension reactions, and determining the pattern of nucleotide
incorporation.

Example 1; Fig 2; 86pp; English.

CC The invention relates to a novel method for obtaining typing information
CC about several variable sites within target nucleic acid, or typing one or
CC more nucleic acid molecules. The methods of the invention are useful for
CC typing one or more nucleic acid molecules containing two or more variable
CC sites, preferably nucleic acid molecules containing three or more
CC variable sites are typed, where three or more primer extension reactions
CC are performed. The method is also useful for diagnosis of pathological
CC conditions characterized by the presence of specific nucleic acid
CC molecules. The methods are particularly suited for identifying
CC microbial species or their subtypes, and in typing procedures e.g. typing
CC of polymorphisms, tissue typing or in clinical applications. The sequence
CC represents the 5' untranslated region (UTR) of a hepatitis C virus (HCV)
CC genotype, amplified in the invention to type HCV-positive sera.

SQ Sequence 305 BP; 56 A; 87 C; 98 G; 64 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 241; DB 24; Length 305;

Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGGCTAGCAGCCTAGTATGATGATGCTGAGCCTCCAGACCCCT 60
DB 26 GCAGAAAGGCTAGCAGCCTAGTATGATGATGCTGAGCCTCCAGACCCCT 85
QY 61 CCCGGAGAGCCATAGTGTGTGCGGAACCGGTGATGATGCTGAGCCTCCAGACCC 120
DB 86 CCCGGAGAGCCATAGTGTGTGCGGAACCGGTGATGATGCTGAGCCTCCAGACCC 145
QY 121 GGGTCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 146 GGGTCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 205
QY 181 TGGTACCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 206 TGGTACCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 265
QY 241 C 241
DB 266 C 266

RESULT 2

ABL46275 standard; RNA; 328 BP.

XX ABL46275;

DT 26-APR-2002 (first entry)

DE Hepatitis C virus subtype 1a RNA sequence SEQ ID NO:242.

KW Nucleic acid accessible hybridisation site; detection; hybridisation;
KW characterisation; identification; nucleic acid structure; diagnosis;
KW gene; ss.

OS Hepatitis C virus.

XX MO200198537-A2.

PD 27-DEC-2001.

PF 15-JUN-2001; 2001WO-US19401.

PR 17-JUN-2000; 2000US-212308.

XX 15-JUN-2001; 2001US-0212308.

PA (THIR-) THIRD WAVE TECHNOLOGIES INC.

XX
PI Lyamichev V, Allawi H, Dong F, Nei BP, Vener IT;
XX
DR WPI; 2002-049698/06.

PT Identifying oligonucleotides hybridizing to nucleic acids containing
PT secondary structure, useful in clinical diagnosis, comprises
PT identifying primers that interact with the target to form an extension
PT product under amplification conditions.

PS Claim 48; Fig 74; 409pp; English.

CC The present invention describes a method for identifying oligonucleotides
CC with desired hybridisation properties to nucleic acid targets containing
CC secondary structure. The method comprises amplifying a target nucleic
CC acid having at least one accessible and one inaccessible site. Primers
CC that form an extension product are identified as the oligonucleotides
CC which can interact with the folded target nucleic acid. Oligonucleotides
CC from the present invention can be used in novel detection methods for
CC clinical diagnostic purposes, including the detection and identification
CC of pathogenic organisms (e.g. HIV). The method allows the ability to
CC rapidly analyse nucleic acid structures. ABL46034 to ABL46367 represent
CC sequences used in the exemplification of the present invention.

SQ Sequence 328 BP; 63 A; 95 C; 102 G; 68 U; 0 other;

Query Match

Best Local Similarity 100.0%; Score 241; DB 24; Length 328;

Matches 190; Conservative 51; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGGCTAGCAGCCTAGTATGATGATGCTGAGCCTCCAGACCCCT 60
DB 50 GCAGAAAGGCTAGCAGCCTAGTATGATGATGCTGAGCCTCCAGACCCCT 109
QY 61 CCCGGAGAGCCATAGTGTGTGCGGAACCGGTGATGATGCTGAGCCTCCAGACCC 120
DB 110 CCCGGAGAGCCATAGTGTGTGCGGAACCGGTGATGATGCTGAGCCTCCAGACCC 169
QY 121 GGGTCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 170 GGGTCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 229
QY 181 TGGTACCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 230 TGGTACCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 289
QY 241 C 241
DB 290 C 290

RESULT 3

AAZ08979 standard; DNA; 341 BP.

XX AAZ08979;

DT 02-NOV-1999 (first entry)

DE Hepatitis C virus 5' UTR used as promoter for a HCV genetic vaccine.

KW Genetic vaccine; immunisation; humoral response; cellular response;
KW non-structural protein; NS protein; antigen; ds.

OS Hepatitis C virus.

XX WO9938880-A1.

PD 05-AUG-1999.

PF 28-JAN-1999; 99WO-US01823.

XX 30-JAN-1998; 98US-0073156.

XX (GEHO) GEN HOSPITAL CORP.
PA Encke J, Mands J;
PI WPI: 1999-494077/41.

XX New hepatitis virus nucleic acids for, e.g. inducing an immune
PT response against the virus
PS Claim 6, Page 10-11: 41pp: English.

XX This sequence represents the hepatitis C virus (HCV) 5' untranslated
CC region (5' UTR). This is preferred for use as a promoter in a
CC recombinant DNA vaccine against hepatitis C. The genetic
CC vaccine comprises regulatory elements capable of directing
CC expression in humans and one or more HCV non-structural
CC (NS) proteins. The HCV genome encodes three NS proteins: NS3, a serine
CC protease/viral helicase; NS4, and NS5, the viral RNA-dependent RNA
CC polymerase. In HCV infection, these and other virally encoded proteins
CC are produced by proteolytic processing of a precursor polypeptide.
CC However, for use in a genetic vaccine, the sequences coding for NS
CC proteins had to be engineered so that each NS sequence had its own
CC initiation and stop codons. Restriction sites were also engineered
CC into the gene fragments to aid subcloning. The gene fragments were
CC isolated and mutated using PCR. The NS3 coding sequence was modified
CC using PCR primers AA208980 and AA208981, the NS4 coding sequence
CC modified using AA208984 and AA208985, and the NS5 coding sequence
CC modified using AA208986 and AA208987. Studies in mice immunised with
CC constructs expressing one NS protein were found to elicit strong
CC antigen-specific immune responses in both arms of the immune system. This
CC demonstrated that the NS proteins are better antigens for stimulating
CC humoral immune responses as compared with previous studies using the HCV
CC core structural protein. The genetic vaccine is useful for inducing
CC an immune response (cellular or humoral) against hepatitis C virus in
CC a human uninfected by the virus, and for immunising a human susceptible
CC to hepatitis C viral infection by inducing an immune response. The
CC composition is also useful for treating a human infected with hepatitis
CC C virus, by induction of an immune response. The advantage of this
CC method of immunisation compared with immunisations with soluble
CC recombinant proteins or peptides, is the ability to induce a strong
CC inflammatory CD4+ T cell response as well as cytotoxic T cell activity.
CC In addition, the new recombinant genetic vaccine is more suitable for
CC immunisation, unlike synthetic peptides which only have a limited
CC number of epitopes available for stimulation of the host response.

XX Sequence 341 BP; 63 A; 103 C; 106 G; 69 T; 0 other;

Query Match 100.0%; Score 241; DB 20; Length 341;
Best Local Similarity 100.0%; Pred. No. 3e-62;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGGCTTATGATGATGCTGTCGAGCCCTCCAGAGACCCCT 60
DB 68 GCAGAAAGCGTCTAGCCATGCGGCTTATGATGATGCTGTCGAGCCCTCCAGAGACCCCT 127
QY 61 CCCGGAGAGCCATAGTGTCTGCGGAACCGGTAGTACACCGGAATTGCCAGACGACC 120
DB 128 CCCGGAGAGCCATAGTGTCTGCGGAACCGGTAGTACACCGGAATTGCCAGACGACC 187
QY 121 GGGTCCTTTCTTGTGATCAACCCGCTCATGCTGAGATTGGGCGTGCCTCCCGAGAC 180
DB 188 GGGTCCTTTCTTGTGATCAACCCGCTCATGCTGAGATTGGGCGTGCCTCCCGAGAC 247
QY 181 TGTAGCCGAGTACTGTTGGTGCAGAAAGCGCTTGTGTAATGCTGATAGAGGTGCTTG 240
DB 248 TGTAGCCGAGTACTGTTGGTGCAGAAAGCGCTTGTGTAATGCTGATAGAGGTGCTTG 307
QY 241 C 241
DB 308 C 308

RESULT 4
AA24834

ID AA24834 standard; DNA; 341 BP.

AC AA24834;

DT 21-JUN-1999 (first entry)

DE Infectious hepatitis C virus genotype 1a/1b 5'UTR consensus.

KW HCV; infectious clone; infection; diagnosis; therapy; vaccine;
screening; assay; antiviral; virucide; ss.

OS Hepatitis C virus.

FN W09904008-A2.

PD 28-JAN-1999.

PT 16-JUL-1998; 98WO-0514688.

PR 27-JAN-1998; 98US-0014416.

PR 18-JUL-1997; 97US-0053062.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Bukh J, Emerson SU, Purcell RH, Yanagi M;

WPI: 1999-132252/11.

New isolated hepatitis C virus nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of HCV
PT infections and for developing screening assays

XX Example 7; Fig 11; 126pp: English.

XX The present sequence comprises a consensus of the 5' untranslated
CC regions (5'UTR) of hepatitis C virus (HCV) infectious clones of
CC genotype 1a (PCV-H77C, see AA24836) and 1b (PCV-J4L65, see AA24835).
CC The invention discloses nucleic acid sequences (see AA24832-33 and
CC AA24843) which encode infectious HCV viruses, and the use of these
CC sequences, and polypeptides (see AA08020-22) encoded by them, in the
CC development of vaccines and diagnostics for HCV and in the
CC development of screening assays for the identification of antiviral
CC agents for HCV.

XX Sequence 341 BP; 63 A; 103 C; 106 G; 69 T; 0 other;

Query Match 100.0%; Score 241; DB 20; Length 341;
Best Local Similarity 100.0%; Pred. No. 3e-62;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGGCTTATGATGATGCTGTCGAGCCCTCCAGAGACCCCT 60
DB 68 GCAGAAAGCGTCTAGCCATGCGGCTTATGATGATGCTGTCGAGCCCTCCAGAGACCCCT 127
QY 61 CCCGGAGAGCCATAGTGTCTGCGGAACCGGTAGTACACCGGAATTGCCAGACGACC 120
DB 128 CCCGGAGAGCCATAGTGTCTGCGGAACCGGTAGTACACCGGAATTGCCAGACGACC 187
QY 121 GGGTCCTTTCTTGTGATCAACCCGCTCATGCTGAGATTGGGCGTGCCTCCCGAGAC 180
DB 188 GGGTCCTTTCTTGTGATCAACCCGCTCATGCTGAGATTGGGCGTGCCTCCCGAGAC 247
QY 181 TGTAGCCGAGTACTGTTGGTGCAGAAAGCGCTTGTGTAATGCTGATAGAGGTGCTTG 240
DB 248 TGTAGCCGAGTACTGTTGGTGCAGAAAGCGCTTGTGTAATGCTGATAGAGGTGCTTG 307
QY 241 C 241
DB 308 C 308

RESULT 5

AA24835 standard; DNA; 341 BP.

AA24835:

21-JUN-1999 (first entry)

Infectious hepatitis C virus genotype 1b 5'UTR.

HCV; infectious clone; infection; diagnosis; therapy; vaccine; screening; assay; antiviral; virucide; ss.

Hepatitis C virus.

WO9904008-A2.

28-JAN-1999.

16-JUL-1998; 98WO-US14688.

27-JAN-1998; 98US-0014416.

18-JUL-1997; 97US-0053062.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Bukh J, Emerson SU, Purcell RH, Yanagi M;

WPI; 1999-13252/11.

New isolated hepatitis C virus nucleic acids - used to develop products for the diagnosis, prevention and treatment of HCV infections and for developing screening assays

Example 7; Fig 11; 126pp; English.

The present sequence comprises the 5' untranslated region (5'UTR) of hepatitis C virus (HCV) genotype 1b infectious clone PCV-J4L6S. This has been compared with the 5'UTR of genotype 1a infectious clone PCV-H77C (see AAX24836) and a consensus (see AAX24834) produced. The invention discloses nucleic acid sequences (see AAX24832-33 and AAX24843) which encode infectious HCV viruses, and the use of these sequences, and polypeptides (see AAW98020-22) encoded by them, in the development of vaccines and diagnostics for HCV and in the development of screening assays for the identification of antiviral agents for HCV.

SQ Sequence 341 BP; 63 A; 103 C; 106 G; 69 T; 0 other;

Query Match 100.0%; Score 241; DB 20; Length 341;
 Best Local Similarity 100.0%; Pred. No. 3e-62;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GCAGAAAGCGCTAGCATGCGCTTGTAGTGTGTCACGCTCCAGAGACCCCTT 60
 68 GCAGAAAGCGCTAGCATGCGCTTGTAGTGTGTCACGCTCCAGAGACCCCTT 127
 128 CCGGGAGAGCATTATGCTGTGCGAACCAGGAGTACACCGGAATTCACAGAGAC 187
 121 GGGTCCTTCTTGTGATCAACCCGCTCAATGCTGAGATTTGGGGTGCCTCCGAGAC 180
 188 GGGTCCTTCTTGTGATCAACCCGCTCAATGCTGAGATTTGGGGTGCCTCCGAGAC 247
 181 TGGTACCGAGTAGTGTGGTGGCGAAGGCTTGTGCTACTGCTGATAGGAGTCTTG 240
 248 TGGTACCGAGTAGTGTGGTGGCGAAGGCTTGTGCTACTGCTGATAGGAGTCTTG 307
 241 C 241
 308 C 308

RESULT 6

AA25773/C standard; DNA; 342 BP.

AA25773:

05-APR-2000 (first entry)

Hepatitis C virus antisense inhibitor oligonucleotide #39.

Hepatitis C virus; HCV; antisense oligonucleotide; hepatotropic; ss; anti-inflammatory; translation inhibition; HCV infection; virucide.

Hepatitis C virus.

US6001990-A.

14-DEC-1999.

07-JUN-1995; 95US-0474700.

10-MAY-1994; 94US-0240382.

(GENO) GEN HOSPITAL CORP.

Moradpour D, Wands JR, Wakita T;

WPI; 2000-104900/09.

Antisense oligonucleotide to Hepatitis C virus RNA, useful for treating Hepatitis C virus infections

Claim 28; Column 31; 31pp; English.

This sequence is an antisense oligonucleotide that hybridizes to Hepatitis C virus (HCV) RNA, under physiological conditions. The invention relates to HCV antisense oligonucleotides, and also for a vector comprising a nucleotide sequence, which is transcribed in an animal cell to generate an antisense oligonucleotide. The oligonucleotides have virucide, hepatotropic and anti-inflammatory activity, and are useful for treating HCV infection by inhibiting translation of type I-V HCV RNA. Hepatitis C virus is a positive strand RNA virus, and is the major causative agent of post-transfusion hepatitis. Persistent HCV infection can lead to chronic hepatitis, cirrhosis, and hepatocellular carcinoma.

SQ Sequence 342 BP; 69 A; 106 C; 103 G; 64 T; 0 other;

Query Match 100.0%; Score 241; DB 21; Length 342;
 Best Local Similarity 100.0%; Pred. No. 3e-62;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GCAGAAAGCGCTAGCATGCGCTTGTAGTGTGTCACGCTCCAGAGACCCCTT 60
 275 GCAGAAAGCGCTAGCATGCGCTTGTAGTGTGTCACGCTCCAGAGACCCCTT 216
 61 CCGGGAGAGCATTATGCTGTGCGAACCAGGAGTACACCGGAATTCACAGAGAC 120
 215 CCGGGAGAGCATTATGCTGTGCGAACCAGGAGTACACCGGAATTCACAGAGAC 156
 121 GGGTCCTTCTTGTGATCAACCCGCTCAATGCTGAGATTTGGGGTGCCTCCGAGAC 180
 155 GGGTCCTTCTTGTGATCAACCCGCTCAATGCTGAGATTTGGGGTGCCTCCGAGAC 96
 181 TGGTACCGAGTAGTGTGGTGGCGAAGGCTTGTGCTACTGCTGATAGGAGTCTTG 240
 95 TGGTACCGAGTAGTGTGGTGGCGAAGGCTTGTGCTACTGCTGATAGGAGTCTTG 36
 241 C 241
 35 C 35

DF	15-MAR-2002	(first entry)	
XX	Hepatitis C virus RNA polymerase encoding polynucleotide SEQ ID NO 4.		
XX	Hepatitis C virus; RNA polymerase; RNA virus replication; ds.		
XX	Hepatitis C virus.		
OS	Hepatitis C virus.		
XX	MO20018161-A1.		
PN	22-NOV-2001.		
XX	15-MAY-2001; 2001WO-JP04033.		
XX	15-MAY-2000; 2000JP-0142451.		
PR	(TOKM-) TOKYO METROPOLITAN ORG MEDICAL RES.		
PA	(CHUS) CHUGAI SEIYAKU KK.		
XX	Kohara M, Matsuzaki J, Okamoto K, Katsune T;		
XX	WPI: 2002-114236/15.		
DR	Vector for analysing RNA virus replication, contains DNA encoding RNA		
PT	virus-derived RNA dependent RNA polymerase		
XX	Example 1; Page 55; 66pp; Japanese.		
XX	The invention relates to a vector for analysing RNA virus replication,		
CC	comprising DNA encoding RNA virus-derived RNA dependent RNA polymerase		
CC	(ABA96556-ABA96281), for evaluation of RNA virus-derived RNA dependent		
CC	RNA polymerase activity by studying interaction between viral proteins,		
CC	host factors and the level of expression of reporter genes.		
SO	Sequence 412 BP; 92 A; 119 C; 120 G; 81 T; 0 other;		
Query Match	100.0%; Score 241; DB 24; Length 412;		
Best Local Similarity	100.0%; Pred. No. 3.1e-62;		
Matches 241; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 GCAGAAACCGCTGACCATGGCGCTTAGTATGAGTGTCTGTGAGCGCTCCAGAACCCCCCT 60		
DB	99 GCAGAAACCGCTGACCATGGCGCTTAGTATGAGTGTCTGTGAGCGCTCCAGAACCCCCCT 158		
OY	61 CCGGGAGAGCGCATGATGCGTGTGCGGAACCGGTGAGTACACCGGAATTGCCAGAGCACC 120		
DB	159 CCGGGAGAGCGCATGATGCGTGTGCGGAACCGGTGAGTACACCGGAATTGCCAGAGCACC 218		
OY	121 GGGTCCTTCTTGATCAACCCGCTCAATGCTTGAGATTGGCGCGTCCCGCCGAGAC 180		
DB	219 GGGTCCTTCTTGATCAACCCGCTCAATGCTTGAGATTGGCGCGTCCCGCCGAGAC 278		
OY	181 TGCTAGCCGAGTAGTGGTGGGCGGAAGGCGCTTGATGCTGCTGATGAGGTCCTTG 240		
DB	279 TGCTAGCCGAGTAGTGGTGGGCGGAAGGCGCTTGATGCTGCTGATGAGGTCCTTG 338		
OY	241 C 241		
DB	339 C 339		
RESULT 9			
AA032444			
ID	AA032444 standard; DNA; 483 BP.		
AC	AA032444;		
XX	26-APR-1993 (first entry)		
XX	HCV core-envelope clone NI-1.		
XX	Clone; polypeptide; core-envelope; region; Hepatitis C; Virus; HCV;		
KW	HC; cDNA; transcriptase; primer; allele; diagnostic; antibody;		

KM suppress; control; proteolytic processing; precursor; ss.
 XX Hepatitis C virus.
 XX
 XX Key Location/Qualifiers
 FT CDS 318..482
 FT /*tag= a
 XX
 PN EP518313-A.
 PD 16-DEC-1992.
 XX
 PF 11-JUN-1992; 92EP-0109812.
 XX
 PR 11-JUN-1991; 91JP-0139268.
 PR 12-JUL-1991; 91JP-0172794.
 PR 07-OCT-1991; 91JP-0287008.
 PR 16-DEC-1991; 91JP-0332329.
 PR 20-APR-1992; 92JP-0099957.
 XX
 PA (MITU) MITSUBISHI KASEI CORP.
 PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
 PI Teranishi Y;
 XX
 DR WPI: 1992-417213/51.
 DR P-PDB: AAR29534.
 XX
 PT New hepatitis C virus gene and its encoded protein - used for
 PT diagnosing and vaccinating against hepatitis C virus infections
 XX
 PS Claim 33; Page 74; 305pp; English.
 XX
 CC The sequences given in AA032438-39 and AA032444-53 are various clones
 CC which encode the polypeptide core-envelope region of the Hepatitis C
 CC virus (HCV) gene of the invention. These sequences were isolated from
 CC the serum of a patient suffering from hepatitis C (HC). The isolated
 CC RNA sequences were converted into cDNA using transcriptase in the
 CC presence of one of the primer sequences given in AA032540-46. The cDNA
 CC sequences isolated represent different alleles of the same region of
 CC the HCV gene. The entire HCV gene (see AA032436) is useful in the
 CC development of a diagnostic method which is more accurate and
 CC effective than conventional ones. In the detection of antibodies
 CC raised against a wide range of HCVs which have been hardly detected
 CC before. The complete gene may be used in an in vitro screening system
 CC for a substance capable of specifically suppressing or controlling a
 CC proteolytic processing of a precursor polypeptide of HCV.
 CC
 XX
 SO Sequence 483 BP; 102 A; 141 C; 146 G; 94 T; 0 other;
 Query Match 100.0%; Score 241; DB 13; Length 483;
 Best Local Similarity 100.0%; Pred. No. 3.2e-62;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGCTAGCCATGGCGTATGATGATGCTGTCAGACCTCCAGAGACCCCT 60
 Db 44 GCAGAAAGCGCTAGCCATGGCGTATGATGATGCTGTCAGACCTCCAGAGACCCCT 103
 Oy 61 CCGGAGAGAGCCATGATGCTGTCAGACCTCCAGAGATTAACCGGAAATTCAGAGAC 120
 Db 104 CCGGAGAGAGCCATGATGCTGTCAGACCTCCAGAGATTAACCGGAAATTCAGAGAC 163
 Oy 121 GGGTCTTTCTTGATCAACCCGCTCAATGCTGAGATTTGGGGTGGCCCGGAGAAC 180
 Db 164 GGGTCTTTCTTGATCAACCCGCTCAATGCTGAGATTTGGGGTGGCCCGGAGAAC 223
 Oy 181 TCGTAGCCGATAGTGTGGTGGTGGAGAAAGGCTTGTGATGCTGATAGGAGTGTG 240
 Db 224 TCGTAGCCGATAGTGTGGTGGTGGAGAAAGGCTTGTGATGCTGATAGGAGTGTG 283
 Oy 241 C 241
 Db 284 C 284

RESULT 10
 AA032445
 ID AA032445 standard; DNA; 483 BP.
 XX
 AC AA032445;
 XX
 DT 26-APR-1993 (first entry)
 XX
 DE HCV core-envelope clone N1-2.
 XX
 KM Clone; polypeptide; core-envelope; region; Hepatitis C Virus; HCV;
 KM HC; cDNA; transcriptase; primer; allele; diagnostic; antibody;
 KM suppress; control; proteolytic processing; precursor; ss.
 OS Hepatitis C virus.
 XX
 XX Key Location/Qualifiers
 FT CDS 318..482
 FT /*tag= a
 XX
 PN EP518313-A.
 PD 16-DEC-1992.
 XX
 PF 11-JUN-1992; 92EP-0109812.
 XX
 PR 11-JUN-1991; 91JP-0139268.
 PR 12-JUL-1991; 91JP-0172794.
 PR 07-OCT-1991; 91JP-0287008.
 PR 16-DEC-1991; 91JP-0332329.
 PR 20-APR-1992; 92JP-0099957.
 XX
 PA (MITU) MITSUBISHI KASEI CORP.
 PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
 PI Teranishi Y;
 XX
 DR WPI: 1992-417213/51.
 DR P-PDB: AAR29535.
 XX
 PT New hepatitis C virus gene and its encoded protein - used for
 PT diagnosing and vaccinating against hepatitis C virus infections
 XX
 PS Claim 33; Page 87; 305pp; English.
 XX
 CC The sequences given in AA032438-39 and AA032444-53 are various clones
 CC which encode the polypeptide core-envelope region of the Hepatitis C
 CC virus (HCV) gene of the invention. These sequences were isolated from
 CC the serum of a patient suffering from hepatitis C (HC). The isolated
 CC RNA sequences were converted into cDNA using transcriptase in the
 CC presence of one of the primer sequences given in AA032540-46. The cDNA
 CC sequences isolated represent different alleles of the same region of
 CC the HCV gene. The entire HCV gene (see AA032436) is useful in the
 CC development of a diagnostic method which is more accurate and
 CC effective than conventional ones. In the detection of antibodies
 CC raised against a wide range of HCVs which have been hardly detected
 CC before. The complete gene may be used in an in vitro screening system
 CC for a substance capable of specifically suppressing or controlling a
 CC proteolytic processing of a precursor polypeptide of HCV.
 CC
 XX
 SO Sequence 483 BP; 102 A; 141 C; 146 G; 94 T; 0 other;
 Query Match 100.0%; Score 241; DB 13; Length 483;
 Best Local Similarity 100.0%; Pred. No. 3.2e-62;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGCTAGCCATGGCGTATGATGATGCTGTCAGACCTCCAGAGACCCCT 60
 Db 44 GCAGAAAGCGCTAGCCATGGCGTATGATGATGCTGTCAGACCTCCAGAGACCCCT 103
 Oy 61 CCGGAGAGAGCCATGATGCTGTCAGACCTCCAGAGATTAACCGGAAATTCAGAGAC 120

Db 104 CCGGGAGAGACCATAGTGTGCTGGAGAACGGGTGATGACACCGAATTGCGAGAGACACC 163
|||
Qy 121 GGGTCTTTTGTGGATCACCCTCATCTGAGATTGGGCGTGCCTCCCGAGAC 180
|||
Db 164 GGGTCTTTTGTGGATCACCCTCATCTGAGATTGGGCGTGCCTCCCGAGAC 223
|||
Qy 181 TGCTAGCCGAGTGTGTGGTGGCGGAAAGCCCTGTGTGATGCTGATAGGTTGCTTG 240
|||
Db 224 TGCTAGCCGAGTGTGTGGTGGCGGAAAGCCCTGTGTGATGCTGATAGGTTGCTTG 283
|||
Qy 241 C 241
|||
Db 284 C 284

RESULT 11

AAZ57395 standard; DNA; 587 BP.

XX ID AAZ57395 standard; DNA; 587 BP.
XX AC AAZ57395;
XX DF 07-APR-2000 (first entry)
XX DE Hepatitis C virus DNA sequence SEQ ID NO:2.
XX KM Hepatitis C virus; RNA virus; replication; viral infection; ds.
XX OS Hepatitis C virus.
XX OS Hepatitis C virus.
XX PN WO967394-A1.
XX PD 29-DEC-1999.
XX PF 24-JUN-1999; 99WO-JP03380.
XX PR 24-JUN-1998; 98JP-0177820.
XX PA (CHUS) CHUGAI SEIRAKU KK.
XX PI Kohara M, Kohara K, Taira K, Matsuzaki J, Ohmori H;
XX DR WPI; 2000-106296/09.
XX PT Vectors expressing full-length gene of RNA viruses, useful in
XX PT clarifying mechanisms of RNA viral replication, infection, and
XX PT developing remedies and therapeutics
XX PS Example 1; Page 35; 46pp; Japanese.
XX CC The present invention describes a vector comprising a cDNA encoding an
XX CC RNA virus gene, constructed to ensure the exact and homogeneous
XX CC transcription of both terminals of the RNA virus gene. Also described
XX CC is a method for screening drugs for inhibiting the replication of RNA
XX CC virus by using the RNA viral infection model animal, particularly one
XX CC with hepatitis C viral infection. The vector is useful in clarifying
XX CC the mechanism of RNA viral infection, onset of RNA viral infection,
XX CC and developing remedies and therapeutics for RNA viral infections,
XX CC particularly of a hepatitis C virus. The present sequence represents
XX CC a DNA sequence from hepatitis C virus from an example of the present
XX CC invention.
XX SQ Sequence 587 BP; 122 A; 178 C; 178 G; 109 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 241; DB 21; Length 587;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAGAAAGCGCTACGATGCGCTGATGAGTGTGCTGAGCTTCCAGAGACCCCTT 60
|||
Db 42 GCAGAAAGCGCTACGATGCGCTGATGAGTGTGCTGAGCTTCCAGAGACCCCTT 101
|||
Qy 61 CCGGAGAGACCATAGTGTGCTGGAGAACGGGTGATGACACCGAATTGCGAGAGAC 120

Db 102 CCGGGAGAGACCATAGTGTGCTGGAGAACGGGTGATGACACCGAATTGCGAGAGACACC 161
|||
Qy 121 GGGTCTTTTGTGGATCACCCTCATCTGAGATTGGGCGTGCCTCCCGAGAC 180
|||
Db 162 GGGTCTTTTGTGGATCACCCTCATCTGAGATTGGGCGTGCCTCCCGAGAC 221
|||
Qy 181 TGCTAGCCGAGTGTGTGGTGGCGGAAAGCCCTGTGTGATGCTGATAGGTTGCTTG 240
|||
Db 222 TGCTAGCCGAGTGTGTGGTGGCGGAAAGCCCTGTGTGATGCTGATAGGTTGCTTG 281
|||
Qy 241 C 241
|||
Db 282 C 282

RESULT 12

AAQ44921 standard; DNA; 686 BP.

XX ID AAQ44921 standard; DNA; 686 BP.
XX AC AAQ44921;
XX DF 04-OCT-1994 (first entry)
XX DE Hepatitis C virus 5'-untranslated region and part of core region.
XX KM Hepatitis C virus; HCV, non-A, non-B hepatitis virus; NANBH;
XX KM antisense oligonucleotide; translation inhibition; therapy;
XX KM 5'-untranslated region; 5'-UTR; hybridisation target; ss.
XX OS Hepatitis C virus.
XX OS Hepatitis C virus.
XX FH Key Location/Qualifiers
XX FT 1..341
XX FT 5'UTR /*tag= a
XX CDS 342..686 /*tag= b
XX FT /product= Core protein
XX FT /note= "partial coding region"
XX PA (KAGA) CHERO SERO THERAPEUTIC RES INST.
XX PA (ISIS-) ISIS PHARM INC.
XX PA (MOCH) MOCHIDA PHARM CO LTD.
XX PI Anderson KP, Eto T, Furukawa S, Hamada F, Hanecak RC;
XX PI Hoshiko K, Nakatake H, Nishihara T, Nozaki C;
XX DR WPI; 1994-101217/12.
XX PT Anti-sense oligo:nucleotide(s) complementary to hepatitis C viral
XX PT genome - useful for inhibiting HCV replication, to treat related
XX PT diseases
XX PS Disclosure; Fig 1; 91pp; English.
XX CC Oligonucleotides which are complementary to part of the hepatitis C
XX CC virus genomic or messenger RNA are claimed. Preferred antisense
XX CC oligonucleotides are complementary to RNA comprising regions of the
XX CC 5'-UTR, esp. the 5'end hairpin loop, 5'end 6bp repeat and 5'end
XX CC untranslated region.

Query Match Sequence 686 BP; 129 A; 214 C; 221 G; 1 T; 0 other;

Best Local Similarity 100.0%; Score 241; DB 15; Length 686;

Best Local Similarity 79.3%; Pred. No. 3.4e-62;
Matches 191; Conservative 50; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTTCAGGCTAGGCGTTCAGTGTGCTGTCGACGCTCCAGACCCCT 60
DB 68 GCAGAAAGCGTTCAGGCTAGGCGTTCAGTGTGCTGTCGACGCTCCAGACCCCT 127
QY 61 CCCGGAGAGCCATAGTGTGTCGCGAACCCTGATACACCGGAATTCACGAGAC 120
DB 128 CCCGGAGAGCCATAGTGTGTCGCGAACCCTGATACACCGGAATTCACGAGAC 187
QY 121 GGGTCTTCTTGTGATCAACCCCTCATGCTGAGATTGGGGCTGCCCGGAGAC 180
DB 188 GGGTCTTCTTGTGATCAACCCCTCATGCTGAGATTGGGGCTGCCCGGAGAC 247
QY 181 TGCCTAGCCGAGTGTGTCGCGAACCCTGATACACCGGAATTCACGAGAC 240
DB 248 TGCCTAGCCGAGTGTGTCGCGAACCCTGATACACCGGAATTCACGAGAC 307
QY 241 C 241
DB 308 C 308

RESULT 13

AAZ57396
ID AAZ57396 standard; DNA; 703 BP.

XX AC AAZ57396;
XX DT 07-APR-2000 (first entry)
XX DE Hepatitis C virus DNA sequence SEQ ID NO:3.
XX DE Hepatitis C virus; RNA virus; replication; viral infection; ds.
XX OS Hepatitis C virus.
XX PN MO9967394-A1.
XX PD 29-DEC-1999.
XX PF 24-JUN-1999; 99MO-JP03380.
XX PR 24-JUN-1998; 98JP-0177820.
XX PA (CHUS) CHUGAI SEIRYAKU KK.
XX PI Kohara M, Kohara K, Taira K, Matsuzaki J, Ohmori H;
XX DR WPI; 2000-106296/09.

PT Vectors expressing full-length gene of RNA viruses, useful in
PT clarifying mechanisms of RNA viral replication, infection, and
PT developing remedies and therapeutics

PS Example 1; Page 35-36; 46pp; Japanese.

CC The present invention describes a vector comprising a cDNA encoding an
CC RNA virus gene, constructed to ensure the exact and homogeneous
CC transcription of both terminals of the RNA virus gene. Also described
CC is a method for screening drugs for inhibiting the replication of RNA
CC virus by using the RNA viral infection model animal, particularly one
CC with hepatitis C viral infection. The vector is useful in clarifying
CC the mechanism of RNA viral replication, onset of RNA viral infection,
CC and developing remedies and therapeutics for RNA viral infection,
CC particularly of a hepatitis C virus. The present sequence represents
CC a DNA sequence from hepatitis C virus from an example of the present
CC invention.

Sequence 703 BP; 151 A; 211 C; 213 G; 128 T; 0 other;

Query Match

100.0%; Score 241; DB 21; Length 703;

Best Local Similarity 100.0%; Pred. No. 3.4e-62;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTTCAGGCTAGGCGTTCAGTGTGCTGTCGACGCTCCAGACCCCT 60
DB 158 GCAGAAAGCGTTCAGGCTAGGCGTTCAGTGTGCTGTCGACGCTCCAGACCCCT 217
QY 61 CCCGGAGAGCCATAGTGTGTCGCGAACCCTGATACACCGGAATTCACGAGAC 120
DB 218 CCCGGAGAGCCATAGTGTGTCGCGAACCCTGATACACCGGAATTCACGAGAC 277
QY 121 GGGTCTTCTTGTGATCAACCCCTCATGCTGAGATTGGGGCTGCCCGGAGAC 180
DB 278 GGGTCTTCTTGTGATCAACCCCTCATGCTGAGATTGGGGCTGCCCGGAGAC 337
QY 181 TGCCTAGCCGAGTGTGTCGCGAACCCTGATACACCGGAATTCACGAGAC 240
DB 338 TGCCTAGCCGAGTGTGTCGCGAACCCTGATACACCGGAATTCACGAGAC 397
QY 241 C 241
DB 398 C 398

RESULT 14

AAA08097
ID AAA08097 standard; cDNA; 713 BP.

XX AC AAA08097;
XX DT 22-JUN-2000 (first entry)
XX DE Hepatitis type C virus nucleotide sequence SEQ ID NO:1.
XX DE Hepatitis C virus; HCV; 5' UTR; 5'-nontranslational region; diagnosis;
XX KW gene expression; infection; IRES; viral; ss.
XX OS Hepatitis C virus.

XX FH Key
XX FT 5'UTR
XX FT 1..341
XX FT /*tag= a
XX FT 342..713
XX FT /*tag= b
XX FT /note= "no stop codon given"
XX PN MO200012691-A1.

XX PD 09-MAR-2000.
XX PF 08-JUL-1999; 99MO-JP03682.
XX PR 27-AUG-1998; 98JP-0241367.
XX PA (FUSO) FUSO PHARM IND LTD.
XX PI Yamada O, Yoshida H, Zhang J;
XX DR WPI; 2000-237867/20.
XX DR P-PSDB; AAY82358.

PT Nucleic acid sequence which potentiates gene expression and contains
PT 5'-non translated region of a viral genome is incorporated in vectors
PT for improving the expression of a useful gene

PS Claim 18; Page 82-84; 94pp; Japanese.

CC The present invention describes an oligonucleotide sequence which
CC potentiates the expression of a useful gene when incorporated in a
CC gene expression vector, where the oligonucleotide sequence consists
CC of the 5'-non translated region (5'UTR) of a viral genome or its
CC fragment or modified form. The 5'UTR sequence is useful in screening
CC potential initiation factors of interaction with the viral IRES or

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 285.951 seconds
(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631B-3

Perfect score: 33
Sequence: 1 cgggtactacacgtccgcagacacatgac 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: GenBank:*
2: gb_ba:*
3: gb_htg:*
4: gb_in:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	30.4	92.1	36	6	A52660	A52660 Sequence 5
C 2	27.4	83.0	29	6	A68291	A68291 Sequence 12
C 3	22	66.7	68	6	AX003355	AX003355 Sequence
C 4	22	66.7	68	6	AX003356	AX003356 Sequence
C 5	22	66.7	116	14	HPCCHA12	M74254 Hepatitis C
C 6	22	66.7	116	14	HPCCHA13	M74255 Hepatitis C
C 7	22	66.7	125	14	HPCCHA10	M74252 Hepatitis C
C 8	22	66.7	140	6	E10301	E10301 Anti-sense
C 9	22	66.7	142	14	S72378	S72378 (5' region)
C 10	22	66.7	148	14	HPCCHA6	M74248 Hepatitis C
C 11	22	66.7	148	14	S75838	S75838 NS-5 (5' un
C 12	22	66.7	149	14	HPCEB1	D10114 Hepatitis C
C 13	22	66.7	149	14	HPCEB11	D10119 Hepatitis C
C 14	22	66.7	149	14	HPCEB12	D10119 Hepatitis C
C 15	22	66.7	149	14	HPCEB13	D10120 Hepatitis C
C 16	22	66.7	149	14	HPCEB16	D10121 Hepatitis C
C 17	22	66.7	149	14	HPCEB18	D10122 Hepatitis C
C 18	22	66.7	149	14	HPCEB2	D10113 Hepatitis C
C 19	22	66.7	149	14	HPCEB7	D10115 Hepatitis C
C 20	22	66.7	149	14	HPCEB8	D10116 Hepatitis C
C 21	22	66.7	149	14	HPCEB9	D10117 Hepatitis C
C 22	22	66.7	155	14	HC056523	U56523 Hepatitis C
C 23	22	66.7	155	14	HC056524	U56524 Hepatitis C
C 24	22	66.7	155	14	HC056525	U56525 Hepatitis C
C 25	22	66.7	155	14	HC056526	U56526 Hepatitis C
C 26	22	66.7	155	14	HC056527	U56527 Hepatitis C
C 27	22	66.7	155	14	HC056528	U56528 Hepatitis C
C 28	22	66.7	155	14	HC056529	U56529 Hepatitis C
C 29	22	66.7	155	14	HC056530	U56530 Hepatitis C
C 30	22	66.7	155	14	HC056531	U56531 Hepatitis C
C 31	22	66.7	155	14	HC056532	U56532 Hepatitis C
C 32	22	66.7	155	14	HC056533	U56533 Hepatitis C
C 33	22	66.7	155	14	HC056534	U56534 Hepatitis C
C 34	22	66.7	155	14	HC056535	U56535 Hepatitis C
C 35	22	66.7	155	14	HC056536	U56536 Hepatitis C
C 36	22	66.7	155	14	HC056537	U56537 Hepatitis C
C 37	22	66.7	155	14	HC056538	U56538 Hepatitis C
C 38	22	66.7	155	14	HC056539	U56539 Hepatitis C
C 39	22	66.7	155	14	HC056540	U56540 Hepatitis C
C 40	22	66.7	155	14	HC056541	U56541 Hepatitis C
C 41	22	66.7	155	14	HC056542	U56542 Hepatitis C
C 42	22	66.7	155	14	HC056543	U56543 Hepatitis C
C 43	22	66.7	155	14	HC056544	U56544 Hepatitis C
C 44	22	66.7	155	14	HC056545	U56545 Hepatitis C
C 45	22	66.7	155	14	HC056546	U56546 Hepatitis C

ALIGNMENTS

RESULT 1
LOCUS A52660 36 bp DNA 11near PAT 12-DEC-1997
DEFINITION Sequence 5 from Patent WO9624662.
ACCESSION A52660
VERSION A52660.1 GI:2851824
KEYWORDS
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 36)
AUTHORS Ravagnan G., Battaglia M., Carloni G., Ponzetto A. and Iacovacci S.
TITLE PROCESS TO 'IN VITRO' PROPAGATE THE HEPATITIS C VIRUS (HCV) IN NON
LYMPHOBLASTOID ANIMAL CELL CULTURES AND PRODUCTS THEREOF
JOURNAL Patent: WO 9624662-A 5 15-AUG-1996;

COMMENT
CONSIGLIO NAZIONALE RICERCA (IT)
Other publication AU 1822195 960827.
FEATURES
Location/Qualifiers
1. 36
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT
9 a 9 c 11 g 7 t
ORIGIN

Query Match
Best Local Similarity 92.1%; Score 30.4; DB 6; Length 36;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y
1 CGGTGTTACTACCGTTCCGACGACCACTATGG 32
32 CGGTGTTACTACCGTTCCGACGACCACTATGG 1

RESULT 2
A68291/c
LOCUS A68291. Sequence 12 from Patent WO9746716. 29 bp DNA linear PAT 06-MAY-1999
DEFINITION
ACCESSION A68291
VERSION A68291.1 GI:4759412
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.

REFERENCE
AUTHORS Bosio, P., Strumia, C. and Clementza, F.
TITLE METHOD TO DETECT HCV SPECIFIC NUCLEIC ACIDS
JOURNAL Patent: WO 9746716-A 12 11-DEC-1997;
WABCO B V. (NL)
Other publication IF RM960404 19971209.

COMMENT
FEATURES
Location/Qualifiers
1. 29
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT
7 a 7 c 9 g 6 t
ORIGIN

Query Match
Best Local Similarity 83.0%; Score 27.4; DB 6; Length 29;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y
4 TGTACTACCGTTCCGACGACCACTATGG 32
29 TGTACTACCGTTCCGACGACCACTATGG 1

RESULT 3
AX003355/c
LOCUS AX003355. Sequence 6 from Patent WO9928503. 68 bp DNA linear PAT 24-AUG-2000
DEFINITION
ACCESSION AX003355
VERSION AX003355.1 GI:9927160
KEYWORDS
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.

REFERENCE
AUTHORS Primi, D. and Mantero, G.
TITLE Methods of detecting polynucleotide analytes
JOURNAL Patent: WO 9928503-A 6 10-JUN-1999;
PRIMI DANIELE (IT); MANTERO GIOVANNI (IT)
Location/Qualifiers
1. 68
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

FEATURES
source
14 a 17 c 24 g 13 t
BASE COUNT
ORIGIN

Query Match
Best Local Similarity 66.7%; Score 22; DB 6; Length 68;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Y
1 CGGTGTTACTACCGTTCCGACGACCACTATGGC 33
40 CGGTGTTACTACCGTTCCGACGACCACTATGGC 7

RESULT 4
AX003356
LOCUS AX003356. Sequence 7 from Patent WO9928503. 68 bp DNA linear PAT 24-AUG-2000
DEFINITION
ACCESSION AX003356
VERSION AX003356.1 GI:9927161
KEYWORDS
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.

REFERENCE
AUTHORS Primi, D. and Mantero, G.
TITLE Methods of detecting polynucleotide analytes
JOURNAL Patent: WO 9928503-A 7 10-JUN-1999;
PRIMI DANIELE (IT); MANTERO GIOVANNI (IT)
Location/Qualifiers
1. 68
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT
13 a 24 c 17 g 14 t
ORIGIN

Query Match
Best Local Similarity 66.7%; Score 22; DB 6; Length 68;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Y
1 CGGTGTTACTACCGTTCCGACGACCACTATGGC 33
29 CGGTGTTACTACCGTTCCGACGACCACTATGGC 62

RESULT 5
HPCCHA12/c
LOCUS HPCCHA12. 116 bp ss-RNA linear VRL 02-AUG-1993
DEFINITION
ACCESSION M74254
VERSION M74254.1 GI:329743
KEYWORDS
SOURCE Hepatitis C virus cDNA to genomic RNA.
ORGANISM Hepatitis C virus

REFERENCE
AUTHORS Cha, T.-A., Kolberg, J., Irvine, B., Stempelen, M., Beall, E., Yano, M., Choo, Q.-L., Houghton, M., Kuo, G., Han, J. H. and Ueda, M. S.
TITLE Use of a signature nucleotide sequence of the hepatitis C virus for the detection of viral RNA in human serum and plasma
JOURNAL Unpublished (1991)
Location/Qualifiers
1. 116
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT
23 a 40 c 33 g 20 t
ORIGIN
Map position 108.

Query Match
Best Local Similarity 66.7%; Score 22; DB 14; Length 116;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Y
1 CGGTGTTACTACCGTTCCGACGACCACTATGGC 33
63 CGGTGTTACTACCGTTCCGACGACCACTATGGC 30

RESULT 6
LOCUS HPCCHA13 116 bp ss-RNA linear VRL 02-AUG-1993
DEFINITION Hepatitis C virus isolate 1128 genome, 5' untranslated region.
ACCESSION M74255
VERSION M74255.1 GI:329744
KEYWORDS
SOURCE Hepatitis C virus cDNA to genomic RNA.
ORGANISM Hepatitis C virus
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE 1 (bases 1 to 116)
AUTHORS Cha,T.-A., Kolberg,J., Irvine,B., Stempien,M., Beall,E., Yano,M.,
Choo,Q.-L., Houghton,M., Kuo,G., Han,J.H. and Urdea,M.S.
TITLE Use of a signature nucleotide sequence of the hepatitis C virus for
the detection of viral RNA in human serum and plasma
JOURNAL Unpublished (1991)
FEATURES
source Location/Qualifiers
1..116
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 23 a 40 c 33 g 20 t
ORIGIN Map position 108.
Query Match 66.7%; Score 22; DB 14; Length 116;
Best Local Similarity 97.1%; Pred. No. 4.8e+02;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 1 CGGTGTACTACAC-GTTCGCGAGACCACTATGCG 33
Db 63 CGGTGTACTACACCGGTTCGCGAGACCACTATGCG 30
RESULT 7
LOCUS HPCCHA10/c 125 bp ss-RNA linear VRL 02-AUG-1993
DEFINITION Hepatitis C virus isolate 1124 genome, 5' untranslated region.
ACCESSION M74252
VERSION M74252.1 GI:329741
KEYWORDS
SOURCE Hepatitis C virus cDNA to genomic RNA.
ORGANISM Hepatitis C virus
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE 1 (bases 1 to 125)
AUTHORS Cha,T.-A., Kolberg,J., Irvine,B., Stempien,M., Beall,E., Yano,M.,
Choo,Q.-L., Houghton,M., Kuo,G., Han,J.H. and Urdea,M.S.
TITLE Use of a signature nucleotide sequence of the hepatitis C virus for
the detection of viral RNA in human serum and plasma
JOURNAL Unpublished (1991)
FEATURES
source Location/Qualifiers
1..125
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 24 a 42 c 37 g 22 t
ORIGIN Map position 99.
Query Match 66.7%; Score 22; DB 14; Length 125;
Best Local Similarity 97.1%; Pred. No. 4.8e+02;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 1 CGGTGTACTACAC-GTTCGCGAGACCACTATGCG 33
Db 72 CGGTGTACTACACCGGTTCGCGAGACCACTATGCG 39
RESULT 8
LOCUS E10301 140 bp RNA linear PAT 29-SEP-1997
DEFINITION Anti-sense RNA against partial sequence of 5'UTR of hepatitis C
virus.
ACCESSION E10301

VERSION E10301.1 GI:5708588
KEYWORDS JP 1995303485-A/2.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 140)
AUTHORS Funahashi,S. and Hasegawa,A.
TITLE HCV ANTI-SENSE RNA, MANIFESTATION VECTOR CONTAINING THE SAME, AND
THERAPEUTIC AGENT FOR HCV-INVOLVED DISEASE CONTAINING THE RNA OR
MANIFESTATION VECTOR
JOURNAL Patent: JP 1995303485-A 2 21-NOV-1995;
TONEIN CORP
COMMENT OS None
OC Artificial sequences.
PN JP 1995303485-A/2
PD 21-NOV-1995
PE 13-MAY-1994 JP 1994124609
PI FUNAHASHI SHINICHI, HASEGAWA AKIRA
PC C12N15/09,A61K31/70,A61K31/70,C07H21/04;
CC strandedness: Single;
CC topology: linear;
CC hypothetical: No;
CC anti-sense: Yes;
FH key Location/Qualifiers
FH source 1..140
FT /organism="Artificial sequences" FT
FT misc-feature 1..140 /note="Anti-sense RNA against partial sequence
of 5'UTR of
hepatitis C virus".
FT Location/Qualifiers
1..140
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 26 a 43 c 46 g 25 t
ORIGIN
Query Match 66.7%; Score 22; DB 6; Length 140;
Best Local Similarity 97.1%; Pred. No. 4.7e+02;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 1 CGGTGTACTACAC-GTTCGCGAGACCACTATGCG 33
Db 80 CGGTGTACTACACCGGTTCGCGAGACCACTATGCG 113
RESULT 9
LOCUS S72378/c 142 bp RNA linear VRL 24-JAN-1995
DEFINITION [5' region] [hepatitis C virus HCV, host-human liver, Genomic RNA,
142 nt].
ACCESSION S72378
VERSION S72378.1 GI:632885
KEYWORDS
SOURCE Hepatitis C virus host-human liver.
ORGANISM Hepatitis C virus
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE 1 (bases 1 to 142)
AUTHORS Sullivan,D.E. and Gerber,M.A.
TITLE Conservation of hepatitis C virus 5' untranslated sequences in
hepatocellular carcinoma and the surrounding liver
JOURNAL Hepatology 19 (3), 551-553 (1994)
MEDLINE 94164633
PUBMED 8119678
REMARK Genbank staff at the National Library of Medicine created this
entry [NCBI g1bseq 152236] from the original journal article.
This sequence comes from Fig. 1.
FEATURES
source Location/Qualifiers
1..142
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT

30 a 41 c 43 g 28 t

gene

/db_xref="taxon:41856"

1.148

/partial

/gene="NS-5"

Query Match 66.7%; Score 22; DB 14; Length 142;
 Best Local Similarity 97.1%; Pred. No. 4.7e+02;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 33
 |||||||
 106 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 73

RESULT 10

HPCCH6/c

148 bp ss-RNA linear VRL 02-NDG-1993
 Hepatitis C virus isolate NB011 genome, 5' untranslated region.

ACCESSION M74248
 VERSION M74248.1 GI:329749

Hepatitis C virus cDNA to genomic RNA.

SOURCE ORGANISM

Hepatitis C virus positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE 1 (bases 1 to 148)

AUTHORS

TITLE

JOURNAL

FEATURES

BASE COUNT

ORIGIN

Map position 76.

Query Match

Best Local Similarity

Matches

OY

Db

RESULT 11

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REMARK

FEATURES

source

BASE COUNT 29 a 40 c 50 g 29 t

Query Match 66.7%; Score 22; DB 14; Length 148;
 Best Local Similarity 97.1%; Pred. No. 4.7e+02;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 33
 |||||||
 39 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 6

RESULT 12

HPCB1/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REMARK

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

OY

Db

RESULT 13

HPCB1/c

LOCUS

DEFINITION

Hepatitis C virus (HCV) genomic RNA, 5' non-coding region.

HPCB11

HPCB11/c

HPCB11

HPCB11

HPCB11

HPCB11

ACCESSION D10118
VERSION GI:221550
KEYWORDS Hepatitis C virus (isolate:E-b12) CDNA to genomic RNA.
SOURCE Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE
AUTHORS Chan,S.W., Mcomish,F., Holmes,E.C., Dow,B., Peutherer,J.F., Follett,E., Yap,P.L. and Simmonds,P.
TITLE Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants
JOURNAL J. Gen. Virol. 73 (Pt 5), 1131-1141 (1992)
MEDLINE 92268871
REFERENCE 2 (bases 1 to 149)
AUTHORS Chan,S.-W.
TITLE Direct Submission
JOURNAL Submitted (25-DEC-1991) Shiu-Wan Chan, University of Edinburgh, Medical School, Dept. of Medical Microbiology, Teviot Place, Edinburgh EH8 9AG U.K. (E-mail: ECHCASTLE.ED.AC.UK, Tel:031-650-3138, Fax:031-662-4135)
COMMENT Submitted (25-Dec-1991) to DDBJ by: Shiu-Wan Chan
Department of Medical Microbiology
University of Edinburgh
Medical School
Teviot Place
Edinburgh EH8 9AG
U.K.
E mail: ECHCASTLE.ED.AC.UK
Phone: 031-650-3138
Fax: 031-662-4135.

FEATURES
source Location/Qualifiers
1. 149
/organism="Hepatitis C virus"
/isolate="E-b12"
/db_xref="taxon:11103" 31 t

BASE COUNT 29 a 47 c 42 g 31 t

ORIGIN
Query Match 66.7%; Score 22; DB 14; Length 149;
Best Local Similarity 97.1%; Pred. No. 4.7e+02;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGACTCACC-GTTCGCGACGACCTATGCG 33
|||||
74 CGGTGACTCACC-GTTCGCGACGACCTATGCG 41
|||||

RESULT 14
HPCB12/C 149 bp RNA linear VRL 02-FEB-1999
DEFINITION Hepatitis C virus (HCV) genomic RNA, 5' non-coding region.
ACCESSION D10119
VERSION D10119.1 GI:221551
KEYWORDS
SOURCE Hepatitis C virus (isolate:E-b12) CDNA to genomic RNA.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE
AUTHORS Chan,S.W., Mcomish,F., Holmes,E.C., Dow,B., Peutherer,J.F., Follett,E., Yap,P.L. and Simmonds,P.
TITLE Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants
JOURNAL J. Gen. Virol. 73 (Pt 5), 1131-1141 (1992)
MEDLINE 92268871
REFERENCE 2 (bases 1 to 149)
AUTHORS Chan,S.-W.
TITLE Direct Submission
JOURNAL Submitted (25-DEC-1991) Shiu-Wan Chan, University of Edinburgh, Medical School, Dept. of Medical Microbiology, Teviot Place, Edinburgh EH8 9AG U.K. (E-mail: ECHCASTLE.ED.AC.UK,

COMMENT Tel:031-650-3138, Fax:031-662-4135)
Submitted (25-Dec-1991) to DDBJ by:
Shiu-Wan Chan
Department of Medical Microbiology
University of Edinburgh
Medical School
Teviot Place
Edinburgh EH8 9AG
U.K.
E mail: ECHCASTLE.ED.AC.UK
Phone: 031-650-3138
Fax: 031-662-4135.

FEATURES
source Location/Qualifiers
1. 149
/organism="Hepatitis C virus"
/isolate="E-b12"
/db_xref="taxon:11103" 31 t

BASE COUNT 28 a 47 c 43 g 31 t

ORIGIN
Query Match 66.7%; Score 22; DB 14; Length 149;
Best Local Similarity 97.1%; Pred. No. 4.7e+02;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGACTCACC-GTTCGCGACGACCTATGCG 33
|||||
74 CGGTGACTCACC-GTTCGCGACGACCTATGCG 41
|||||

RESULT 15
HPCB13/C 149 bp RNA linear VRL 02-FEB-1999
DEFINITION Hepatitis C virus (HCV) genomic RNA, 5' non-coding region.
ACCESSION D10120
VERSION D10120.1 GI:221554
KEYWORDS
SOURCE Hepatitis C virus (isolate:E-b13) CDNA to genomic RNA.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE
AUTHORS Chan,S.W., Mcomish,F., Holmes,E.C., Dow,B., Peutherer,J.F., Follett,E., Yap,P.L. and Simmonds,P.
TITLE Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants
JOURNAL J. Gen. Virol. 73 (Pt 5), 1131-1141 (1992)
MEDLINE 92268871
REFERENCE 2 (bases 1 to 149)
AUTHORS Chan,S.-W.
TITLE Direct Submission
JOURNAL Submitted (25-DEC-1991) Shiu-Wan Chan, University of Edinburgh, Medical School, Dept. of Medical Microbiology, Teviot Place, Edinburgh EH8 9AG U.K. (E-mail: ECHCASTLE.ED.AC.UK, Tel:031-650-3138, Fax:031-662-4135)
COMMENT Submitted (25-Dec-1991) to DDBJ by: Shiu-Wan Chan
Department of Medical Microbiology
University of Edinburgh
Medical School
Teviot Place
Edinburgh EH8 9AG
U.K.
E mail: ECHCASTLE.ED.AC.UK
Phone: 031-650-3138
Fax: 031-662-4135.

FEATURES
source Location/Qualifiers
1. 149
/organism="Hepatitis C virus"
/isolate="E-b13"
/db_xref="taxon:11103" 28 t

BASE COUNT 27 a 48 c 46 g 28 t

ORIGIN

Query Match 66.7%; Score 22; DB 14; Length 149;
 Best Local Similarity 97.1%; Pred. No. 4.7e+02;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTACACC-GTTCGCGACACCACTATGGC 33
 |||||||||||||
 Db 74 CGGTGTACTACACCGGTTCGCGACACCACTATGGC 41

Search completed: July 10, 2003, 19:20:01
 Job time : 286.951 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 242.625 Seconds

(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631B-7

Perfect score: 28

Sequence: 1 gcaagcaccctatcagcagcaccacaa 28

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl:

1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_yi:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pla:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	28	AR054576	AR054576 Sequence
2	28	100.0	28	AR094138	AR094138 Sequence
3	28	100.0	28	AX147022	AX147022 Sequence
4	28	100.0	47	AR204716	AR204716 Sequence
5	28	100.0	47	AX284180	AX284180 Sequence
6	28	100.0	57	I73305	I73305 Sequence 36
7	28	100.0	75	AX021668	AX021668 Sequence
8	28	100.0	186	E10302	E10302 Anti-sense
9	28	100.0	187	AX172757	AX172757 Sequence
10	28	100.0	191	AF041266	AF041266 Hepatitis
11	28	100.0	191	AF041273	AF041273 Hepatitis
12	28	100.0	191	AF041276	AF041276 Hepatitis
13	28	100.0	191	AF041277	AF041277 Hepatitis
14	28	100.0	191	AF041279	AF041279 Hepatitis
15	28	100.0	191	AF041289	AF041289 Hepatitis
16	28	100.0	191	AF041292	AF041292 Hepatitis
17	28	100.0	191	AF041293	AF041293 Hepatitis
18	28	100.0	191	AF041294	AF041294 Hepatitis
19	28	100.0	191	AF041296	AF041296 Hepatitis
20	28	100.0	191	AF041298	AF041298 Hepatitis
21	28	100.0	191	AF041299	AF041299 Hepatitis
22	28	100.0	191	AF041300	AF041300 Hepatitis
23	28	100.0	191	AF041301	AF041301 Hepatitis
24	28	100.0	191	AF041304	AF041304 Hepatitis
25	28	100.0	191	AF041306	AF041306 Hepatitis
26	28	100.0	191	AF041308	AF041308 Hepatitis
27	28	100.0	191	AF041309	AF041309 Hepatitis
28	28	100.0	191	AF041310	AF041310 Hepatitis
29	28	100.0	191	AF041312	AF041312 Hepatitis
30	28	100.0	191	AF041313	AF041313 Hepatitis
31	28	100.0	197	HCV6315	AJ006315 Hepatitis
32	28	100.0	204	HCV6311	AJ006311 Hepatitis
33	28	100.0	204	HCV6312	AJ006312 Hepatitis
34	28	100.0	205	HCV6313	AJ006313 Hepatitis
35	28	100.0	219	AF506662	AF506662 Hepatitis
36	28	100.0	222	AF387732	AF387732 Hepatitis
37	28	100.0	222	AF387733	AF387733 Hepatitis
38	28	100.0	222	AF506641	AF506641 Hepatitis
39	28	100.0	222	AF506663	AF506663 Hepatitis
40	28	100.0	223	AF506624	AF506624 Hepatitis
41	28	100.0	224	AF506630	AF506630 Hepatitis
42	28	100.0	224	AF506635	AF506635 Hepatitis
43	28	100.0	226	AF506674	AF506674 Hepatitis
44	28	100.0	227	HCV6327	AJ006327 Hepatitis
45	28	100.0	228	AF506654	AF506654 Hepatitis

ALIGNMENTS

RESULT 1
LOCUS AR054576 28 bp DNA
DEFINITION Sequence 2 from patent US 5837442
ACCESSION AR054576
VERSION AR054576.1 GI:5860153
KEYWORDS
SOURCE Unknown
ORGANISM Unknown
REFERENCE 1 (bases 1 to 28)
AUTHORS Tsang,S.Yen.
TITLE Oligonucleotide primers for amplifying HCV nucleic acid
JOURNAL Patent: US 5837442-A 2 17-NOV-1998;
FEATURES Location/Qualifiers

```

SOURCE          1. 28
                /organism="unknown"
BASE COUNT      10 a 10 c 5 g 3 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 28; DB 6; Length 28;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGACACCTATCAGGAGTACCA 28
    1 GCAGACACCTATCAGGAGTACCA 28

RESULT 2
LOCUS AR094138 28 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 4 from patent US 6001611.
ACCESSION AR094138
VERSION AR094138.1 GI:10020883
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 28)
AUTHORS Will, S.Gordon.
TITLE Modified nucleic acid amplification primers
JOURNAL Patent: US 6001611-A 14-DEC-1999;
FEATURES
    Location/Qualifiers
        source
            10 a 10 c 5 g 3 t

Query Match
Best Local Similarity 100.0%; Score 28; DB 6; Length 28;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGACACCTATCAGGAGTACCA 28
    1 GCAGACACCTATCAGGAGTACCA 28

RESULT 3
LOCUS AX147022 28 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 16 from Patent WO0137291.
ACCESSION AX147022
VERSION AX147022.1 GI:14346293
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 28)
AUTHORS Weinidel, K., Riedling, M. and Geiger, A.
TITLE Magnetic glass particles, method for their preparation and uses
JOURNAL Patent: WO 0137291-A 16-25-MAY-2001;
FEATURES
    Roche Diagnostics GmbH (DB)
    Location/Qualifiers
        source
            1. 28
                /organism="synthetic construct"
                /db_xref="taxon:32630"
                /note="Synthetic oligonucleotide (HCV reverse)"
modified_base
    28
    /note="derivation with a p-(t-butyl)benzyl-residue"
    /mod_base=OTHER
BASE COUNT      10 a 10 c 5 g 3 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 28; DB 6; Length 28;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 GCAGACACCTATCAGGAGTACCA 28
    1 GCAGACACCTATCAGGAGTACCA 28

RESULT 4
LOCUS AR204716/c 47 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6368801.
ACCESSION AR204716
VERSION AR204716.1 GI:21502113
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 47)
AUTHORS Faruqi, A.Fawad.
TITLE Detection and amplification of RNA using target-mediated ligation
JOURNAL Patent: US 6368801-A 1 09-APR-2002;
FEATURES
    Location/Qualifiers
        source
            1. 47
                /organism="unknown"
BASE COUNT      7 a 11 c 18 g 11 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 28; DB 6; Length 47;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGACACCTATCAGGAGTACCA 28
    36 GCAGACACCTATCAGGAGTACCA 9

RESULT 5
LOCUS AX284180 47 bp DNA linear PAT 20-NOV-2001
DEFINITION Sequence 1 from Patent WO0179420.
ACCESSION AX284180
VERSION AX284180.1 GI:17044868
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1
AUTHORS Faruqi, A.F.
TITLE Detection and amplification of rna using target-mediated ligation
JOURNAL Patent: WO 0179420-A 1 25-OCT-2001;
FEATURES
    MOLECULAR STAGING, INC. (US)
    Location/Qualifiers
        source
            1. 47
                /organism="synthetic construct"
                /db_xref="taxon:32630"
                /note="Synthetic Target"
BASE COUNT      7 a 11 c 18 g 11 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 28; DB 6; Length 47;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGACACCTATCAGGAGTACCA 28
    36 GCAGACACCTATCAGGAGTACCA 9

RESULT 6
LOCUS I73305 57 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 36 from patent US 5686272.

```

ACCESSION I73305
VERSION I73305.1 GI:3009444
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 57)
TITLE Marshall, R.L., Carrino, J.J. and Sustache, J.C.
JOURNAL Amplification of RNA sequences using the ligase chain reaction
PATENT: US 5686272-A 36 11-NOV-1997;
FEATURES
source 1..57
/organism="unknown"
BASE COUNT 9 a 9 c 23 g 16 t
ORIGIN
Query Match 100.0%; Score 28; DB 6; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAAGCACCCTATCAGGAGTACACAA 28
|||||
54 GCAAGCACCCTATCAGGAGTACACAA 27
RESULT 7
LOCUS AX021668/c 75 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 47 from Patent WO9923250.
ACCESSION AX021668
VERSION AX021668.1 GI:10044951
KEYWORDS
SOURCE
ORGANISM Hepatitis C virus.
Hepatitis C virus.
Virus; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 75)
AUTHORS Kessler, C., Bartl, K., Habershausen, G. and Orum, H.
TITLE Specific and sensitive method for detecting nucleic acids
JOURNAL Patent: WO 9923250-A 47 14-MAY-1999;
KESLER CHRISTOPH (DE); BARTL KNOT (DE); HABERSHAUSEN GERD (DE);
ROCHE DIAGNOSTICS GMBH (DE); OROM HENRIK (DK)
FEATURES
source 1..75
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 13 a 20 c 25 g 17 t
ORIGIN
Query Match 100.0%; Score 28; DB 6; Length 75;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAAGCACCCTATCAGGAGTACACAA 28
|||||
30 GCAAGCACCCTATCAGGAGTACACAA 3
RESULT 8
LOCUS E10302 186 bp RNA linear PAT 29-SEP-1997
DEFINITION Anti-sense RNA against partial sequence of 5'UTR of hepatitis C virus.
ACCESSION E10302
VERSION E10302
KEYWORDS JP 1995303485-A/3.
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 186)
AUTHORS Funahashi, S. and Hasegawa, A.
TITLE HCV ANTI-SENSE RNA, MANIFESTATION VECTOR CONTAINING THE SAME, AND THERAPEUTIC AGENT FOR HCV-INVOLVED DISEASE CONTAINING THE RNA OR

JOURNAL
PATENT: JP 1995303485-A 3 21-NOV-1995;
TONEIN CORP
COMMENT
OS None
OC Artificial sequences.
PN JP 1995303485-A/3
PD 21-NOV-1995
PE 13-MAY-1994 JP 1994124609
PI FUNAHASHI SHINICHI, HASEGAWA AKIRA
PC C12N15/09, A61K31/70, A61K31/70, C07H21/04;
CC strandedness: Single;
CC topology: linear;
CC hypothetical: No;
CC anti-sense: Yes;
FH Key
FH Location/Qualifiers
FT source 1..186
FT /organism="Artificial sequences" FT
FT misc-feature 1..186
FT /note="Anti-sense RNA against partial sequence of 5'UTR of hepatitis C virus".
FT
FEATURES
source 1..186
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 41 a 65 c 48 g 32 t
ORIGIN
Query Match 100.0%; Score 28; DB 6; Length 186;
Best Local Similarity 100.0%; Pred. No. 0.096;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAAGCACCCTATCAGGAGTACACAA 28
|||||
10 GCAAGCACCCTATCAGGAGTACACAA 37
RESULT 9
LOCUS AX172757/c 187 bp mRNA linear PAT 03-JUL-2001
DEFINITION Sequence 5 from Patent WO0144266.
ACCESSION AX172757
VERSION AX172757.1 GI:14597853
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 187)
AUTHORS Karn, J.C. and Walker, S.C.
TITLE Nucleic acid compounds and screening assays using the same
JOURNAL Patent: WO 0144266-A 5 21-JUN-2001;
RiboTargets Limited (GB)
FEATURES
source 1..187
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Probe"
BASE COUNT 35 a 48 c 64 g 40 t
ORIGIN
Query Match 100.0%; Score 28; DB 6; Length 187;
Best Local Similarity 100.0%; Pred. No. 0.096;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAAGCACCCTATCAGGAGTACACAA 28
|||||
178 GCAAGCACCCTATCAGGAGTACACAA 151
RESULT 10
LOCUS AF041266/c 191 bp ss-RNA linear VRL 06-DEC-2000
AF041266

DEFINITION Hepatitis C virus isolate 611 5' untranslated region.
 AF041266
 VERSION AF041266.1 GI:2809125
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

1 (bases 1 to 191)
 Quarleri, J.F., Robertson, B.H., Mathet, V.L., Feld, M., Espinola, L.,
 Requeno, M.P., Mando, O., Carbajal, G. and Oubina, J.R.
 Genomic and phylogenetic analysis of hepatitis C virus isolates
 from Argentine patients: a six-year retrospective study
 J. Clin. Microbiol. 38 (12), 4560-4568 (2000)
 2 (bases 1 to 191)
 Quarleri, J., Robertson, B.H. and Oubina, J.R.
 Direct Submission
 Submitted (07-JAN-1998) Hepatitis Laboratory, Faculty of Medicine,
 University of Buenos Aires, Paraguay 2155 11th, Buenos Aires 1121,
 Argentina

FEATURES
 SOURCE
 location/Qualifiers

1. 191
 /organism="Hepatitis C virus"
 /isolate="611"
 /db_xref="taxon:11103"
 1. 191
 33 a 52 c 65 g 41 t

Query Match
 Best Local Similarity 100.0%; Score 28; DB 14; Length 191;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGGAGTACCA 28
 184 GCAAGCACCCTATCAGGAGTACCA 157

RESULT 11
 AF041273/c
 LOCUS Hepatitis C virus isolate 726 5' untranslated region.
 DEFINITION AF041273
 ACCESSION AF041273
 VERSION AF041273.1 GI:2809132
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

1 (bases 1 to 191)
 Quarleri, J.F., Robertson, B.H., Mathet, V.L., Feld, M., Espinola, L.,
 Requeno, M.P., Mando, O., Carbajal, G. and Oubina, J.R.
 Genomic and phylogenetic analysis of hepatitis C virus isolates
 from Argentine patients: a six-year retrospective study
 J. Clin. Microbiol. 38 (12), 4560-4568 (2000)
 2 (bases 1 to 191)
 Quarleri, J., Robertson, B.H. and Oubina, J.R.
 Direct Submission
 Submitted (07-JAN-1998) Hepatitis Laboratory, Faculty of Medicine,
 University of Buenos Aires, Paraguay 2155 11th, Buenos Aires 1121,
 Argentina

FEATURES
 SOURCE
 location/Qualifiers

1. 191
 /organism="Hepatitis C virus"
 /isolate="726"
 /db_xref="taxon:11103"
 1. 191
 36 a 49 c 65 g 41 t

Query Match
 Best Local Similarity 100.0%; Score 28; DB 14; Length 191;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGGAGTACCA 28
 184 GCAAGCACCCTATCAGGAGTACCA 157

RESULT 12
 AF041276/c
 LOCUS Hepatitis C virus isolate 760 5' untranslated region.
 DEFINITION AF041276
 ACCESSION AF041276
 VERSION AF041276.1 GI:2809135
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

1 (bases 1 to 191)
 Quarleri, J.F., Robertson, B.H., Mathet, V.L., Feld, M., Espinola, L.,
 Requeno, M.P., Mando, O., Carbajal, G. and Oubina, J.R.
 Genomic and phylogenetic analysis of hepatitis C virus isolates
 from Argentine patients: a six-year retrospective study
 J. Clin. Microbiol. 38 (12), 4560-4568 (2000)
 2 (bases 1 to 191)
 Quarleri, J., Robertson, B.H. and Oubina, J.R.
 Direct Submission
 Submitted (07-JAN-1998) Hepatitis Laboratory, Faculty of Medicine,
 University of Buenos Aires, Paraguay 2155 11th, Buenos Aires 1121,
 Argentina

FEATURES
 SOURCE
 location/Qualifiers

1. 191
 /organism="Hepatitis C virus"
 /isolate="760"
 /db_xref="taxon:11103"
 1. 191
 36 a 49 c 65 g 41 t

Query Match
 Best Local Similarity 100.0%; Score 28; DB 14; Length 191;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGGAGTACCA 28
 184 GCAAGCACCCTATCAGGAGTACCA 157

RESULT 13
 AF041277/c
 LOCUS Hepatitis C virus isolate 768 5' untranslated region.
 DEFINITION AF041277
 ACCESSION AF041277
 VERSION AF041277.1 GI:2809136
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

1 (bases 1 to 191)
 Quarleri, J.F., Robertson, B.H., Mathet, V.L., Feld, M., Espinola, L.,
 Requeno, M.P., Mando, O., Carbajal, G. and Oubina, J.R.
 Genomic and phylogenetic analysis of hepatitis C virus isolates
 from Argentine patients: a six-year retrospective study
 J. Clin. Microbiol. 38 (12), 4560-4568 (2000)
 2 (bases 1 to 191)
 Quarleri, J., Robertson, B.H. and Oubina, J.R.
 Direct Submission
 Submitted (07-JAN-1998) Hepatitis Laboratory, Faculty of Medicine,
 University of Buenos Aires, Paraguay 2155 11th, Buenos Aires 1121,
 Argentina

AUTHORS Quarleri, J., Robertson, B.H. and Oubina, J.R.
 TITLE Direct Submission
 JOURNAL Submitted (07-JAN-1998) Hepatitis Laboratory, Faculty of Medicine,
 University of Buenos Aires, Paraguay 2155 11th, Buenos Aires 1121,
 Argentina

FEATURES
 source Location/Qualifiers
 1. 191
 /organism="Hepatitis C virus"
 /isolate="784"
 /db_xref="taxon:11103"
 1. 191

5'UTR 37 a 48 c 64 g 42 t
 BASE COUNT
 ORIGIN

Query Match 100.0%; Score 28; DB 14; Length 191;
 Best Local Similarity 100.0%; Pred. No. 0.096;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCTATCGAGCAGTACCA 28
 ||||||||||||||||||||
 DB 184 GCAAGCACCTATCGAGCAGTACCA 157

RESULT 14
 AF041279/c 191 bp ss-RNA linear VRL 06-DEC-2000
 LOCUS AF041279
 DEFINITION Hepatitis C virus isolate 784 5' untranslated region.
 ACCESSION AF041279
 VERSION AF041279.1 GI:2809138
 KEYWORDS
 SOURCE Hepatitis C virus.
 ORGANISM Hepatitis C virus.
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.

REFERENCE
 1 (bases 1 to 191)
 Quarleri, J.F., Robertson, B.H., Mathet, V.L., Feld, M., Espinoza, L.,
 Requeijo, M.P., Mendo, O., Carbajal, G. and Oubina, J.R.
 Genomic and phylogenetic analysis of hepatitis C virus isolates
 from Argentine patients: a six-year retrospective study
 J. Clin. Microbiol. 38 (12), 4560-4568 (2000)

JOURNAL
 MEDLINE 20553364
 PUBMED 11101596

REFERENCE
 2 (bases 1 to 191)
 Quarleri, J., Robertson, B.H. and Oubina, J.R.
 Direct Submission
 Submitted (07-JAN-1998) Hepatitis Laboratory, Faculty of Medicine,
 University of Buenos Aires, Paraguay 2155 11th, Buenos Aires 1121,
 Argentina

FEATURES
 source Location/Qualifiers
 1. 191
 /organism="Hepatitis C virus"
 /isolate="784"
 /db_xref="taxon:11103"
 1. 191

5'UTR 34 a 51 c 64 g 42 t
 BASE COUNT
 ORIGIN

Query Match 100.0%; Score 28; DB 14; Length 191;
 Best Local Similarity 100.0%; Pred. No. 0.096;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCTATCGAGCAGTACCA 28
 ||||||||||||||||||||
 DB 184 GCAAGCACCTATCGAGCAGTACCA 157

RESULT 15
 AF041289/c 191 bp ss-RNA linear VRL 06-DEC-2000
 LOCUS AF041289
 DEFINITION Hepatitis C virus isolate 818 5' untranslated region.
 ACCESSION AF041289
 VERSION AF041289.1 GI:2809148
 KEYWORDS

SOURCE Hepatitis C virus.
 ORGANISM Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.

REFERENCE
 1 (bases 1 to 191)
 Quarleri, J.F., Robertson, B.H., Mathet, V.L., Feld, M., Espinoza, L.,
 Requeijo, M.P., Mendo, O., Carbajal, G. and Oubina, J.R.
 Genomic and phylogenetic analysis of hepatitis C virus isolates
 from Argentine patients: a six-year retrospective study
 J. Clin. Microbiol. 38 (12), 4560-4568 (2000)

JOURNAL
 MEDLINE 20553364
 PUBMED 11101596

REFERENCE
 2 (bases 1 to 191)
 Quarleri, J., Robertson, B.H. and Oubina, J.R.
 Direct Submission
 Submitted (07-JAN-1998) Hepatitis Laboratory, Faculty of Medicine,
 University of Buenos Aires, Paraguay 2155 11th, Buenos Aires 1121,
 Argentina

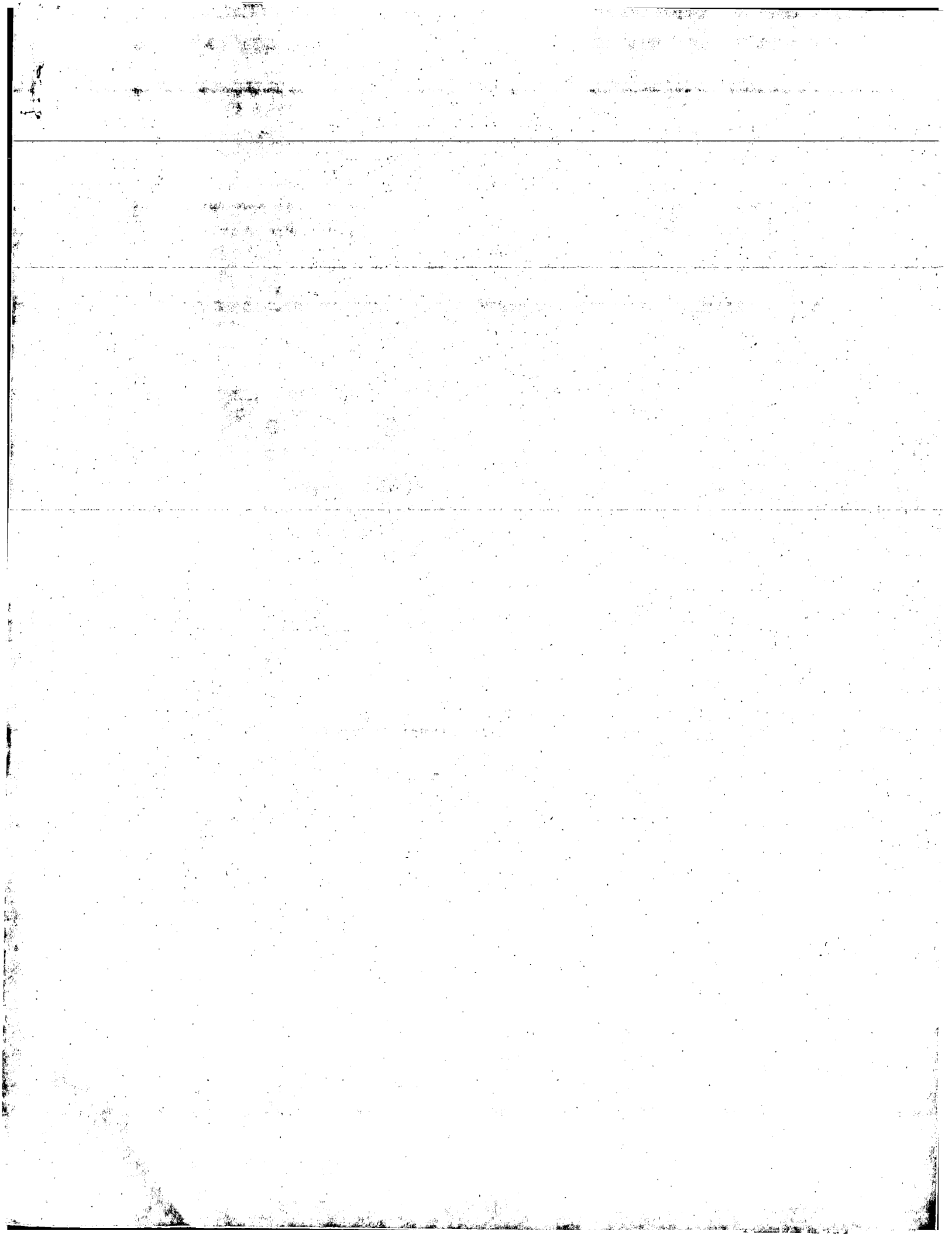
FEATURES
 source Location/Qualifiers
 1. 191
 /organism="Hepatitis C virus"
 /isolate="818"
 /db_xref="taxon:11103"
 1. 191

5'UTR 34 a 51 c 64 g 42 t
 BASE COUNT
 ORIGIN

Query Match 100.0%; Score 28; DB 14; Length 191;
 Best Local Similarity 100.0%; Pred. No. 0.096;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCTATCGAGCAGTACCA 28
 ||||||||||||||||||||
 DB 184 GCAAGCACCTATCGAGCAGTACCA 157

Search completed: July 10, 2003, 19:20:23
 Job time : 243.625 secs



score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

(without alignments)
3358.597 Million cell updates/sec

Sequence: 1 gcagaaagcgtctagccatg'gcgtta 26

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0 ..
Maximum DB seq length: 2000000000
```

Maximum DB seq length: 200000000000

post-processing:	Minimum Match	08
	Maximum Match	100%

Listing first 45 summaries

Database :

```

1:  genbml.*
2:  gp.ba.*
3:  gp.hbg.*
4:  gp.in.*
5:  gp.om.*
6:  gp.ov.*
7:  gp.pat.*
8:  gp.ph.*
9:  gp.pl.*
10: gp.pr.*
11: gp.ro.*
12: gp.sts.*
13: gp.sy.*
14: gp.un.*
15: gp.vi.*
16: em.ba.*
17: em.fun.*
18: em.hum.*
19: em.in.*
20: em.mu.*
21: em.om.*
22: em.or.*
23: em.ov.*
24: em.pat.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.hbg.hum.*
31: em.hbg.in.*
32: em.hbg.inv.*
33: em.hbg.other.*
34: em.hbg.mus.*
35: em.hbg.pln.*
36: em.hbg.rod.*
37: em.hbg.mam.*
38: em.hbg.vrt.*
39: em.hbg.hum.*
40: em.hbg.mus.*
41: em.hbg.other.*

```

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	26	100.0	26	6	AR054575	Sequence
2	26	100.0	26	6	AR094137	Sequence
3	26	100.0	26	6	AX147021	Sequence
4	26	100.0	30	6	AX472307	Sequence
5	26	100.0	51	6	AX021612	Sequence
6	26	100.0	77	6	AX172761	Sequence
7	26	100.0	123	14	HPCUT56CIN	M94468 Hepatitis C
8	26	100.0	137	14	HPCUT55CIN	M94467 Hepatitis C
9	26	100.0	142	14	S72378	S72378 (5' region)
10	26	100.0	155	6	AR095003	Sequence
11	26	100.0	156	14	HPCCHA11	M74253 Hepatitis C
12	26	100.0	160	14	HPCUT74CIN	M94466 Hepatitis C
13	26	100.0	171	14	HPCUT56CIN	M94464 Hepatitis C
14	26	100.0	176	14	HPCUT34CIN	M94461 Hepatitis C
15	26	100.0	176	14	HPCUT38CIN	M94462 Hepatitis C
16	26	100.0	176	14	HPCUTCLN1	M94469 Hepatitis C
17	26	100.0	189	14	AF077227	Hepatitis C
18	26	100.0	190	14	AF021883	AF021883 Hepatitis C
19	26	100.0	190	14	AF021884	AF021884 Hepatitis C
20	26	100.0	190	14	AF021885	AF021885 Hepatitis C
21	26	100.0	190	14	AF021886	AF021886 Hepatitis C
22	26	100.0	190	14	AF021887	AF021887 Hepatitis C
23	26	100.0	190	14	AF021888	AF021888 Hepatitis C
24	26	100.0	190	14	AF021889	AF021889 Hepatitis C
25	26	100.0	190	14	AF021890	AF021890 Hepatitis C
26	26	100.0	190	14	AF021891	AF021891 Hepatitis C
27	26	100.0	190	14	AF021892	AF021892 Hepatitis C
28	26	100.0	190	14	AF021893	AF021893 Hepatitis C
29	26	100.0	190	14	AF021898	AF021898 Hepatitis C
30	26	100.0	190	14	AF021899	AF021899 Hepatitis C
31	26	100.0	190	14	AF021900	AF021900 Hepatitis C
32	26	100.0	190	14	AF021901	AF021901 Hepatitis C
33	26	100.0	190	14	AF021902	AF021902 Hepatitis C
34	26	100.0	190	14	AF021903	AF021903 Hepatitis C
35	26	100.0	190	14	AF021904	AF021904 Hepatitis C
36	26	100.0	194	6	AR066644	Sequence
37	26	100.0	194	6	AR066645	Sequence
38	26	100.0	194	6	AR066646	Sequence
39	26	100.0	194	6	AR066647	Sequence
40	26	100.0	194	6	AR066649	Sequence
41	26	100.0	194	6	AR066650	Sequence
42	26	100.0	194	6	AR066651	Sequence
43	26	100.0	194	6	AR066652	Sequence
44	26	100.0	202	14	HPC5UTR1	D12506 Hepatitis C
45	26	100.0	202	14	HPC5UTR2	D12507 Hepatitis C

ALIGNMENTS

LOCUS	SEQUENCE	LENGTH	REFERENCE
AR054575	Sequence 1 from patent US 5837442.	26 bp	DNA
AR054575	Sequence 1 from patent US 5837442.	1 linear	PAT 29-SEP-1999

ORGANISM

Unclassified.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

```

source
1. 26
BASE COUNT      /organism="unknown"
ORIGIN          7 a      6 c      8 g      5 t

Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 26;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 GCAGAAAGCGCTAGCCATGGCGTTA 26
1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 2
AR094137
LOCUS      AR094137
DEFINITION Sequence 3 from patent US 6001611.
ACCESSION  AR094137
VERSION     AR094137.1  GI:10020882
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 26)
AUTHORS    Will S.Gordon.
TITLE      Modified nucleic acid amplification primers
JOURNAL    Patent: US 6001611-A 3 14-DEC-1999;
FEATURES    Location/Qualifiers
            1..26
            /organism="unknown"
BASE COUNT      7 a      6 c      8 g      5 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 26;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 GCAGAAAGCGCTAGCCATGGCGTTA 26
1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 3
AX147021
LOCUS      AX147021
DEFINITION Sequence 15 from Patent WO0137291.
ACCESSION  AX147021
VERSION     AX147021.1  GI:14346292
KEYWORDS
SOURCE      Synthetic construct;
ORGANISM    synthetic construct;
REFERENCE   1 (bases 1 to 26)
AUTHORS    Weindel, K., Riedling, M. and Geiger, A.
TITLE      Magnetic glass particles; method for their preparation and uses
JOURNAL    Patent: WO 0137291-A 15 25-MAY-2001;
FEATURES    Roche Diagnostics GmbH (DE)
            location/Qualifiers
            1..26
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="Synthetic oligonucleotide primer (HCV forward)"
            /note="derivatization with a p-(t-butyl)benzyl-residue"
            /mod_base=OTHER
BASE COUNT      7 a      6 c      8 g      5 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 26;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY
1 GCAGAAAGCGCTAGCCATGGCGTTA 26
1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 4
AX472307
LOCUS      AX472307
DEFINITION Sequence 17 from Patent WO2052015.
ACCESSION  AX472307
VERSION     AX472307.1  GI:22207326
KEYWORDS
SOURCE      Hepatitis C virus.
ORGANISM    Hepatitis C virus.
REFERENCE   1
AUTHORS    Kukulj, G. and Pause, A.
TITLE      Self-replicating rna molecule from hepatitis C virus
JOURNAL    Patent: WO 02052015-A 17 04-JUL-2002;
BOHRINGER INGENHEIM (CANADA) LTD. (CA)
FEATURES    Location/Qualifiers
            1..30
            /organism="Hepatitis C virus"
            /db_xref="taxon:11103"
BASE COUNT      8 a      7 c      9 g      6 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 30;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 GCAGAAAGCGCTAGCCATGGCGTTA 26
3 GCAGAAAGCGCTAGCCATGGCGTTA 28

RESULT 5
AX021612
LOCUS      AX021612
DEFINITION Sequence 50 from Patent WO9924606.
ACCESSION  AX021612
VERSION     AX021612.1  GI:10044896
KEYWORDS
SOURCE      Hepatitis C virus.
ORGANISM    Hepatitis C virus.
REFERENCE   1 (bases 1 to 51)
AUTHORS    Kessler, C., Bartl, K., Habershausen, G. and Orum, H.
TITLE      Specific and sensitive nucleic acid detection method
JOURNAL    Patent: WO 9924606-A 50 20-MAY-1999;
            KESSLER CHRISTOPH (DE); BARTL KNUT (DE); HABERHAUSEN GERD (DE);
            ROCHE DIAGNOSTICS GMBH (DE); ORUM HENRIK (DK)
FEATURES    Roche Diagnostics GmbH (DE);
            location/Qualifiers
            1..51
            /organism="Hepatitis C virus"
            /db_xref="taxon:11103"
BASE COUNT      11 a      12 c      15 g      13 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 51;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 GCAGAAAGCGCTAGCCATGGCGTTA 26
8 GCAGAAAGCGCTAGCCATGGCGTTA 33

RESULT 6
AX172761

```

LOCUS AX172761 77 bp mRNA linear PAT 03-JUL-2001
 DEFINITION Sequence 9 from Patent WO0144266.
 ACCESSION AX172761
 VERSION AX172761.1 GI:14597857
 KEYWORDS
 ORGANISM
 SOURCE synthetic construct.
 REFERENCE
 AUTHORS Martell,M., Esteban,J.I., Quer,J., Genesca,J., Weiner,A.,
 TITLE Hepatitis C virus (HCV) circulates as a population of different but
 JOURNAL Hepatitis C virus (HCV) circulates as a population of different but
 MEDLINE 92219420
 PUBMED 1313927
 FEATURES
 SOURCE location/Qualifiers
 1..77
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"
 /note="Probe"
 BASE COUNT 16 a 20 c 23 g 18 t
 ORIGIN
 Query Match 100.0%; Score 26; DB 6; Length 77;
 Best Local Similarity 100.0%; Pred. No. 0.054;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 ||||||||||||||||||
 Db 26 GCAGAAAGCGTCTAGCCATGGCGTTA 51

RESULT 7
 HPCUT55CLN 123 bp ss-RNA linear VRL 02-AUG-1993
 LOCUS
 DEFINITION Hepatitis C virus (clone #6) nonstructural protein gene, 5' flank.
 ACCESSION M94468 M84479
 VERSION M94468.1 GI:329981
 KEYWORDS nonstructural protein.
 SOURCE Hepatitis C virus RNA.
 ORGANISM Hepatitis C virus
 REFERENCE
 AUTHORS Martell,M., Esteban,J.I., Quer,J., Genesca,J., Weiner,A.,
 TITLE Hepatitis C virus (HCV) circulates as a population of different but
 JOURNAL Hepatitis C virus (HCV) circulates as a population of different but
 MEDLINE 92219420
 PUBMED 1313927
 FEATURES
 SOURCE location/Qualifiers
 1..123
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"
 BASE COUNT 28 a 35 c 36 g 24 t
 ORIGIN
 Query Match 100.0%; Score 26; DB 14; Length 123;
 Best Local Similarity 100.0%; Pred. No. 0.054;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 ||||||||||||||||||
 Db 23 GCAGAAAGCGTCTAGCCATGGCGTTA 48

RESULT 8
 HPCUT55CLN 137 bp ss-RNA linear VRL 02-AUG-1993
 LOCUS
 DEFINITION Hepatitis C virus (clone #55) nonstructural protein gene, 5' flank.
 ACCESSION M94467.1 GI:329979
 VERSION M94467.1 GI:329979
 KEYWORDS nonstructural protein.

SOURCE Hepatitis C virus RNA.
 ORGANISM Hepatitis C virus
 REFERENCE
 AUTHORS Martell,M., Esteban,J.I., Quer,J., Genesca,J., Weiner,A.,
 TITLE Hepatitis C virus (HCV) circulates as a population of different but
 JOURNAL Hepatitis C virus (HCV) circulates as a population of different but
 MEDLINE 92219420
 PUBMED 1313927
 FEATURES
 SOURCE location/Qualifiers
 1..137
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"
 BASE COUNT 30 a 37 c 41 g 29 t
 ORIGIN
 Query Match 100.0%; Score 26; DB 14; Length 137;
 Best Local Similarity 100.0%; Pred. No. 0.054;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 ||||||||||||||||||
 Db 23 GCAGAAAGCGTCTAGCCATGGCGTTA 48

RESULT 9
 S72378 142 bp RNA linear VRL 24-JAN-1995
 LOCUS
 DEFINITION (5' region) [hepatitis C virus HCV, host-human liver, Genomic RNA,
 142 nt].
 ACCESSION S72378
 VERSION S72378.1 GI:632885
 KEYWORDS
 SOURCE Hepatitis C virus host-human liver.
 ORGANISM Hepatitis C virus
 REFERENCE
 AUTHORS Sullivan,D.E. and Gerber,M.A.
 TITLE Conservation of hepatitis C virus 5' untranslated sequences in
 JOURNAL hepatocellular carcinoma and the surrounding liver
 MEDLINE 94164633
 PUBMED 8119678
 REMARK Genbank staff at the National Library of Medicine created this
 entry [NCBI githsdg 152236] from the original journal article.
 This sequence comes from Fig. 1.
 FEATURES
 SOURCE location/Qualifiers
 1..142
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"
 BASE COUNT 30 a 41 c 43 g 28 t
 ORIGIN
 Query Match 100.0%; Score 26; DB 14; Length 142;
 Best Local Similarity 100.0%; Pred. No. 0.054;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 ||||||||||||||||||
 Db 4 GCAGAAAGCGTCTAGCCATGGCGTTA 29

RESULT 10
 AR095003/c 155 bp DNA linear PAT 08-SEP-2000
 LOCUS
 DEFINITION Sequence 41 from patent US 6001990.
 ACCESSION AR095003
 VERSION AR095003.1 GI:10022459

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
FEATURES

Unknown.
Unknown.
Unclassified.
1 (bases 1 to 155)
Wands, V.R., Wakita, T. and Moradpour, D.
Antisense inhibition of hepatitis C virus
Patent: US 6001990-A 41 14-DEC-1999;
Location/Qualifiers
1..155
/organism="unknown"

BASE COUNT
ORIGIN
29 a 43 c 51 g 32 t

Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 155;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
88 GCAGAAAGCGTCTAGCCATGGCGTTA 63

RESULT 11
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

HPCCH11
Hepatitis C virus isolate 11 genome, 5' untranslated region.
M94253.1 GI:329742
Hepatitis C virus cDNA to genomic RNA.
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C virus

REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE

1 (bases 1 to 156)
Chen, T.-A., Kolberg, J., Irvine, B., Stempien, M., Beall, E., Yano, M.,
Choo, Q.-L., Houghton, M., Kuo, G., Han, J. H. and Urdea, M. S.
Use of a signature nucleotide sequence of the hepatitis C virus for
the detection of viral RNA in human serum and plasma
Unpublished (1991)
Location/Qualifiers
1..156
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT
ORIGIN
31 a 48 c 45 g 32 t
Map position 59.

Query Match
Best Local Similarity 100.0%; Score 26; DB 14; Length 156;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
10 GCAGAAAGCGTCTAGCCATGGCGTTA 35

RESULT 12
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

HPCUT74CLN
Hepatitis C virus (clone #74) nonstructural protein gene, 5' flank.
M94466.1 GI:329982
nonstructural protein.
Hepatitis C virus RNA.
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C virus

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 160)
Martell, M., Esteban, J. I., Quer, J., Genesca, J., Weiner, A.,
Esteban, R., Guardia, J. and Gomez, J.
Hepatitis C virus (HCV) circulates as a population of different but
closely related genomes: quasispaces nature of HCV genome
distribution

JOURNAL
MEDLINE
PUBMED
FEATURES

J. Virol. 66 (5), 3225-3229 (1992)
92219420
1313927
Location/Qualifiers
1..160
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT
ORIGIN
33 a 44 c 48 g 35 t

Query Match
Best Local Similarity 100.0%; Score 26; DB 14; Length 160;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
23 GCAGAAAGCGTCTAGCCATGGCGTTA 48

RESULT 13
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

HPCUT56CLN
Hepatitis C virus (clone #56) nonstructural protein gene, 5' flank.
M94464.1 GI:329980
nonstructural protein.
Hepatitis C virus RNA.
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C virus

REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE

1 (bases 1 to 171)
Martell, M., Esteban, J. I., Quer, J., Genesca, J., Weiner, A.,
Esteban, R., Guardia, J. and Gomez, J.
Hepatitis C virus (HCV) circulates as a population of different but
closely related genomes: quasispaces nature of HCV genome
distribution
J. Virol. 66 (5), 3225-3229 (1992)
92219420
Location/Qualifiers
1..171
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT
ORIGIN
37 a 51 c 47 g 36 t

Query Match
Best Local Similarity 100.0%; Score 26; DB 14; Length 171;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
18 GCAGAAAGCGTCTAGCCATGGCGTTA 43

RESULT 14
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

HPCUT34CLN
Hepatitis C virus (clone #34) nonstructural protein gene, 5' flank.
M94461.1 GI:329976
nonstructural protein.
Hepatitis C virus RNA.
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C virus

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 176)
Martell, M., Esteban, J. I., Quer, J., Genesca, J., Weiner, A.,
Esteban, R., Guardia, J. and Gomez, J.
Hepatitis C virus (HCV) circulates as a population of different but
closely related genomes: quasispaces nature of HCV genome
distribution
J. Virol. 66 (5), 3225-3229 (1992)

MEDLINE 92219420
 PUBMED 1313927
 FEATURES
 source Location/Qualifiers
 1.176

BASE COUNT 37 a 50 c 51 g 38 t
 ORIGIN
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"

Query Match 100.0%; Score 26; DB 14; Length 176;
 Best Local Similarity 100.0%; Pred. No. 0.053;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGCTTA 26
 |||||
 Db 23 GCAGAAAGCGTCTAGCCATGCGCTTA 48

RESULT 15
 HPCUT38CLN 176 bp ss-RNA linear VRL 02-AUG-1993
 LOCUS Hepatitis C virus (clone #38) nonstructural protein gene, 5' flank.
 DEFINITION M94462 M84479
 ACCESSION M94462.1 GI:329977
 VERSION nonstructural protein.
 KEYWORDS Hepatitis C virus RNA.
 SOURCE Hepatitis C virus
 ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.

REFERENCE 1 (bases 1 to 176)
 AUTHORS Martell,M., Esteban,J.I., Quer,J., Genesca,J., Weiner,A.,
 Esteban,R., Guardia,J. and Gomez,J.
 TITLE Hepatitis C virus (HCV) circulates as a population of different but
 closely related genomes: quasispecies nature of HCV genome
 distribution

JOURNAL J. Virol. 66 (5), 3225-3229 (1992)
 MEDLINE 92219420
 PUBMED 1313927

FEATURES
 source Location/Qualifiers
 1.176

BASE COUNT 36 a 51 c 51 g 38 t
 ORIGIN
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"

Query Match 100.0%; Score 26; DB 14; Length 176;
 Best Local Similarity 100.0%; Pred. No. 0.053;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGCTTA 26
 |||||
 Db 23 GCAGAAAGCGTCTAGCCATGCGCTTA 48

Search completed: July 10, 2003, 19:20:22
 Job time : 227.294 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 107.773 Seconds
(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631b-10

Perfect score: 241
Sequence: 1 gcagaagcgtctgcatg.....ctgcctgtaggtgcttgc 241

Scoring table: IDENTITY_NUC
Gapop 10.0 / Gapept 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCtUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/Backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	241	100.0	324	2	US-08-470-426B-15
2	241	100.0	341	3	US-08-854-531-4
3	241	100.0	341	3	US-09-014-416-47
4	241	100.0	341	3	US-09-014-416-48
5	241	100.0	341	4	US-08-869-380-4
6	241	100.0	341	5	PCT-US95-13552-4
7	241	100.0	342	3	US-08-474-700B-39
8	241	100.0	686	4	US-08-988-321B-37
9	241	100.0	686	4	US-08-397-220B-25
10	241	100.0	686	4	US-08-650-093C-25
11	241	100.0	780	3	US-08-474-700B-45
12	241	100.0	923	4	US-08-869-380-1
13	241	100.0	923	4	PCT-US95-13552-14
14	241	100.0	1499	1	US-08-324-977-3
15	241	100.0	1499	2	US-08-384-616-3
16	241	100.0	1499	2	US-08-904-686A-3
17	241	100.0	1499	4	US-09-315-850-3
18	241	100.0	1863	2	US-08-470-426B-14
19	241	100.0	9416	1	US-08-324-977-1
20	241	100.0	9416	2	US-08-384-616-1
21	241	100.0	9416	2	US-08-904-686A-1
22	241	100.0	9416	4	US-09-315-850-1
23	241	100.0	9416	4	US-08-823-885A-27
24	241	100.0	9595	3	US-09-014-416-4
25	241	100.0	9595	3	US-09-014-416-6
26	239.4	99.3	244	4	US-09-034-205-26
27	239.4	99.3	244	4	US-09-034-205-27

28	239.4	99.3	244	4	US-08-934-097A-26	Sequence 26, Appl
29	239.4	99.3	244	4	US-08-934-097A-27	Sequence 27, Appl
30	239.4	99.3	244	4	US-08-851-568-26	Sequence 26, Appl
31	239.4	99.3	244	4	US-08-851-568-27	Sequence 27, Appl
32	239.4	99.3	244	4	US-09-677-218B-26	Sequence 26, Appl
33	239.4	99.3	244	4	US-09-677-218B-27	Sequence 27, Appl
34	239.4	99.3	244	4	US-09-677-192-26	Sequence 26, Appl
35	239.4	99.3	244	4	US-09-677-192-27	Sequence 27, Appl
36	239.4	99.3	281	2	US-08-757-653-123	Sequence 123, App
37	239.4	99.3	281	2	US-08-757-653-128	Sequence 128, App
38	239.4	99.3	281	2	US-08-757-653-129	Sequence 129, App
39	239.4	99.3	281	4	US-08-520-946-123	Sequence 123, App
40	239.4	99.3	281	4	US-08-520-946-128	Sequence 128, App
41	239.4	99.3	281	4	US-08-520-946-129	Sequence 129, App
42	239.4	99.3	286	4	US-09-034-205-21	Sequence 21, Appl
43	239.4	99.3	286	4	US-08-934-097A-21	Sequence 21, Appl
44	239.4	99.3	286	4	US-08-851-568-21	Sequence 21, Appl
45	239.4	99.3	286	4	US-09-677-218B-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-470-426B-15
Sequence 15, Application US/08470426B
Patent No. 5856458
GENERAL INFORMATION:
APPLICANT: Okamoto, Hiroaki
APPLICANT: Nakamura, Tetsuo
TITLE OF INVENTION: OLIGONUCLEOTIDE PRIMERS, AND THEIR
TITLE OF INVENTION: APPLICATION FOR HIGH-FIDELITY DETECTION OF NON-A, NON-B
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Beverlyidge, Degrandi, Wellacher & Young,
L.L.P.
STREET: 1850 M Street, N.W., Suite 800
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,426B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-153402
FILING DATE: 12-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Wellacher, Robert G.
REGISTRATION NUMBER: 20,531
REFERENCE/DOCKET NUMBER: 06/59-47083.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2811
TELEFAX: (202) 659-1462
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-470-426B-15
Query Match 100.0%; Score 241; DB 2; Length 324;
Best Local Similarity 100.0%; Pred. No. 1.5e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGATGTCGTGACGCTCCAGAGACCCCTT 60
Db 51 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGATGTCGTGACGCTCCAGAGACCCCTT 110
QY 61 CCCGGAGAGCCATAGTGTGTCGGGAAACCGGTAGTACACCGGAATTTGCCAGAGAC 120
Db 111 CCCGGAGAGCCATAGTGTGTCGGGAAACCGGTAGTACACCGGAATTTGCCAGAGAC 170
QY 121 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGGAATTTGGCGCTGCCCGAGAGAC 180
Db 171 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGGAATTTGGCGCTGCCCGAGAGAC 230
QY 181 TGTAGCCGAGTATGTTGGGTGCGGAAAGCGCTTGCTGCTAGTATGAGGTCCTTG 240
Db 231 TGTAGCCGAGTATGTTGGGTGCGGAAAGCGCTTGCTGCTAGTATGAGGTCCTTG 290
QY 241 C 241
Db 291 C 291

RESULT 2

US-08-854-531-4
Sequence 4, Application US/08854531
Patent No. 6025341

GENERAL INFORMATION:

APPLICANT: Wands, Jack
APPLICANT: Tokushige, Katsutoshi
TITLE OF INVENTION: CHIMERIC HEPATITIS B/HEPATITIS C VIRUS VACCINE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6025341
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,531
FILING DATE: Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: APOL-0214
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3439
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 341 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-08-854-531-4
Query Match 100.0% Score 241; DB 3; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.6e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGATGTCGTGACGCTCCAGAGACCCCTT 60
Db 68 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGATGTCGTGACGCTCCAGAGACCCCTT 127
QY 61 CCCGGAGAGCCATAGTGTGTCGGGAAACCGGTAGTACACCGGAATTTGCCAGAGAC 120

Db 128 CCCGGAGAGCCATAGTGTGTCGGGAAACCGGTAGTACACCGGAATTTGCCAGAGAC 187
QY 121 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGGAATTTGGCGCTGCCCGAGAGAC 180
Db 188 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGGAATTTGGCGCTGCCCGAGAGAC 247
QY 181 TGTAGCCGAGTATGTTGGGTGCGGAAAGCGCTTGCTGCTAGTATGAGGTCCTTG 240
Db 248 TGTAGCCGAGTATGTTGGGTGCGGAAAGCGCTTGCTGCTAGTATGAGGTCCTTG 307
QY 241 C 241
Db 308 C 308

RESULT 3

US-09-014-416-47
Sequence 47, Application US/09014416
Patent No. 6153421

GENERAL INFORMATION:

APPLICANT: Yanagi, Masayuki
APPLICANT: Burk, Jens
APPLICANT: Emerson, Susanne U.
TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
FILE REFERENCE: 20264276
CURRENT APPLICATION NUMBER: US/09/014,416
EARLIER FILING DATE: 1998-01-27
EARLIER APPLICATION NUMBER: US 60/053,062
EARLIER FILING DATE: 1997-07-18
NUMBER OF SEQ. ID NOS: 65
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 47
LENGTH: 341
TYPE: DNA
ORGANISM: Hepatitis C virus
US-09-014-416-47

Query Match 100.0% Score 241; DB 3; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.6e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGATGTCGTGACGCTCCAGAGACCCCTT 60
Db 68 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGATGTCGTGACGCTCCAGAGACCCCTT 127
QY 61 CCCGGAGAGCCATAGTGTGTCGGGAAACCGGTAGTACACCGGAATTTGCCAGAGAC 120
Db 128 CCCGGAGAGCCATAGTGTGTCGGGAAACCGGTAGTACACCGGAATTTGCCAGAGAC 187
QY 121 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGGAATTTGGCGCTGCCCGAGAGAC 180
Db 188 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGGAATTTGGCGCTGCCCGAGAGAC 247
QY 181 TGTAGCCGAGTATGTTGGGTGCGGAAAGCGCTTGCTGCTAGTATGAGGTCCTTG 240
Db 248 TGTAGCCGAGTATGTTGGGTGCGGAAAGCGCTTGCTGCTAGTATGAGGTCCTTG 307
QY 241 C 241
Db 308 C 308

RESULT 4

US-09-014-416-48
Sequence 48, Application US/09014416
Patent No. 6153421

GENERAL INFORMATION:

APPLICANT: Yanagi, Masayuki
APPLICANT: Burk, Jens
APPLICANT: Emerson, Susanne U.
APPLICANT: Purcell, Robert H.

TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
FILE REFERENCE: 20264276
CURRENT APPLICATION NUMBER: US/09/014,416
CURRENT FILING DATE: 1998-01-27
EARLIER APPLICATION NUMBER: US 60/053,062
EARLIER FILING DATE: 1997-07-18
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 48
LENGTH: 341
TYPE: DNA
ORGANISM: Hepatitis C virus
US-09-014-416-48

Query Match 100.0%; Score 241; DB 3; Length 341;
Best Local Similarity 100.0%; Pred. No. 1,66-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAACGCTCTAGCCATGCGCGTTAGTATGAGTGTGTCGACGCTCCAGAGACCCCT 60
DB 68 GCAGAAACGCTCTAGCCATGCGCGTTAGTATGAGTGTGTCGACGCTCCAGAGACCCCT 127
QY 61 CCCGGAGAGCCATAGTGTGTCGCGAACCGGTGAGTACACCGGAATTGCCAGAGAC 120
DB 128 CCCGGAGAGCCATAGTGTGTCGCGAACCGGTGAGTACACCGGAATTGCCAGAGAC 187
QY 121 GGGTCCTTCTTGATCAACCCGCTCAATGCCGTGAGATTGGCGCTGCCCGGAGAC 180
DB 188 GGGTCCTTCTTGATCAACCCGCTCAATGCCGTGAGATTGGCGCTGCCCGGAGAC 247
QY 181 TGCTAGCCGAGTAGTGTGCGGAAAGCGCTGTGCTACTGCTATAGGCGCTTG 240
DB 248 TGCTAGCCGAGTAGTGTGCGGAAAGCGCTGTGCTACTGCTATAGGCGCTTG 307
QY 241 C 241
DB 308 C 308

RESULT 5

US-08-869-380-4
Sequence 4, Application US/08869380
Patent No. 6235888
GENERAL INFORMATION:
APPLICANT: Pachuk, Catherine J.
APPLICANT: Wands, Jack
APPLICANT: Wakita, Takaji
APPLICANT: Zurawski, Jr., Vincent R.
APPLICANT: Coney, Leslie R.
TITLE OF INVENTION: Hepatitis C Virus Vaccine
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
ADDRESS: No. 6235888ris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/869,380
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/318,248
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: APOL-0189
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 341 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-869-380-4

Query Match 100.0%; Score 241; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 1,66-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAACGCTCTAGCCATGCGCGTTAGTATGAGTGTGTCGACGCTCCAGAGACCCCT 60
DB 68 GCAGAAACGCTCTAGCCATGCGCGTTAGTATGAGTGTGTCGACGCTCCAGAGACCCCT 127
QY 61 CCCGGAGAGCCATAGTGTGTCGCGAACCGGTGAGTACACCGGAATTGCCAGAGAC 120
DB 128 CCCGGAGAGCCATAGTGTGTCGCGAACCGGTGAGTACACCGGAATTGCCAGAGAC 187
QY 121 GGGTCCTTCTTGATCAACCCGCTCAATGCCGTGAGATTGGCGCTGCCCGGAGAC 180
DB 188 GGGTCCTTCTTGATCAACCCGCTCAATGCCGTGAGATTGGCGCTGCCCGGAGAC 247
QY 181 TGCTAGCCGAGTAGTGTGCGGAAAGCGCTGTGCTACTGCTATAGGCGCTTG 240
DB 248 TGCTAGCCGAGTAGTGTGCGGAAAGCGCTGTGCTACTGCTATAGGCGCTTG 307
QY 241 C 241
DB 308 C 308

RESULT 6

PCT-US95-13552-4
Sequence 4, Application PC/TUS9513552
GENERAL INFORMATION:
APPLICANT: Wands, Jack
APPLICANT: Tokushige, Katsutoshi
APPLICANT: Wakita, Takaji
APPLICANT: Pachuk, Catherine J.
APPLICANT: Zurawski, Jr., Vincent R.
APPLICANT: Coney, Leslie R.
TITLE OF INVENTION: HEPATITIS VIRUS VACCINES
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
ADDRESS: Norris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13552
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/318,248
FILING DATE: 05-OCT-1994
CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/467,859
 FILING DATE: 06-JUN-1995
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Deluca, Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: APOL-0238
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 341 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 PCT-US95-13552-4

Query Match 100.0%; Score 241; DB 5; Length 341;
 Best Local Similarity 100.0%; Pred. No. 1.6e-68;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGCGGCTTAGTGTGAGTGTGTCGACGCTCCAGAGACCCCTT 60
 DB 68 GCAGAAAGCGTCTAGCCATGCGGCTTAGTGTGAGTGTGTCGACGCTCCAGAGACCCCTT 127
 OY 61 CCCGGAGAGCCATAGTGTGTCGCGAACCAGGTCAGTACACCGGAATTGCCAGAGACGCC 120
 DB 128 CCCGGAGAGCCATAGTGTGTCGCGAACCAGGTCAGTACACCGGAATTGCCAGAGACGCC 187
 OY 121 GGGTCCTTCTTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 180
 DB 188 GGGTCCTTCTTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 247
 OY 181 TGTACCGAGTAGTGTGTCGCGAAGGCTTGTGATGCTGCTGATGAGGTCCTG 240
 DB 248 TGTACCGAGTAGTGTGTCGCGAAGGCTTGTGATGCTGCTGATGAGGTCCTG 307
 OY 241 C 241
 DB 308 C 308

RESULT 7

US-08-474-700B-39/C
 Sequence 39, Application US/08474700B
 Patent No. 6001990

GENERAL INFORMATION:

APPLICANT: Wands, Jack
 APPLICANT: Moradpour, Darius
 TITLE OF INVENTION: ANTISENSE INHIBITION OF HEPATITIS C
 TITLE OF INVENTION: VIRUS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB
 COMPUTER: IBM PS/2 Model 50Z or 55SX
 OPERATING SYSTEM: MS-DOS (Version 5.0)
 SOFTWARE: Wordperfect (Version 5.1)
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/474,700B
 FILING DATE: 07-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/240,382
 FILING DATE: 10 May 1994

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.
 REGISTRATION NUMBER: 34,819
 REFERENCE/DOCKET NUMBER: 00786/279001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 542-5070
 TELEFAX: (617) 542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 39:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 342 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 US-08-474-700B-39

Query Match 100.0%; Score 241; DB 3; Length 342;
 Best Local Similarity 100.0%; Pred. No. 1.6e-68;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGCGGCTTAGTGTGAGTGTGTCGACGCTCCAGAGACCCCTT 60
 DB 275 GCAGAAAGCGTCTAGCCATGCGGCTTAGTGTGAGTGTGTCGACGCTCCAGAGACCCCTT 216
 OY 61 CCCGGAGAGCCATAGTGTGTCGCGAACCAGGTCAGTACACCGGAATTGCCAGAGACGCC 120
 DB 215 CCCGGAGAGCCATAGTGTGTCGCGAACCAGGTCAGTACACCGGAATTGCCAGAGACGCC 156
 OY 121 GGGTCCTTCTTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 180
 DB 155 GGGTCCTTCTTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 96
 OY 181 TGTACCGAGTAGTGTGTCGCGAAGGCTTGTGATGCTGCTGATGAGGTCCTG 240
 DB 95 TGTACCGAGTAGTGTGTCGCGAAGGCTTGTGATGCTGCTGATGAGGTCCTG 36
 OY 241 C 241
 DB 35 C 35

RESULT 8

US-08-988-321B-37
 Sequence 37, Application US/08988321B
 Patent No. 6174868

GENERAL INFORMATION:

APPLICANT: Kevin P. Anderson et al.
 TITLE OF INVENTION: Compositions And Methods For Treatment Of Hepatitis C V
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Law Offices of Jane Massey Licata
 STREET: 66 East Main Street
 CITY: Marlton
 STATE: NJ
 COUNTRY: USA
 ZIP: 08053

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
 COMPUTER: IBM COMPATIBLE
 OPERATING SYSTEM: WINDOWS 95
 SOFTWARE: WORDPERFECT 6.1 FOR WINDOWS
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/988,321B
 FILING DATE: December 10, 1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/650,093
 FILING DATE: May 17, 1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/452,841
 FILING DATE: May 30, 1995
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/397,220
FILING DATE: March 9, 1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 07/945,289
FILING DATE: September 10, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: ISPH-0245
TELECOMMUNICATION INFORMATION:
TELEPHONE: (856) 810-1515
TELEFAX: (856) 810-1454
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 686
TYPE: nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: No
US-08-988-321B-37

Query Match 100.0%; Score 241; DB 4; Length 686;
Best Local Similarity 79.3%; Pred. No. 1.9e-68;
Matches 191; Conservative 50; Mismatches 0; Indels 0; Gaps 0;

1 GCAGAAAGCGTGTAGCCATGCGGTAGTATGATGCTGCGAGCCCTCCAGAGACCCCT 60
68 GCAGAAAGCGTGTAGCCATGCGGTAGTATGATGCTGCGAGCCCTCCAGAGACCCCT 127
61 CCCGGAGAGCCATAGTGTGCGGAACCGGTGAGTACACCGGAATTGCCAGAGAC 120
128 CCCGGAGAGCCATAGTGTGCGGAACCGGTGAGTACACCGGAATTGCCAGAGAC 187
121 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 180
188 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 247
181 TGCTAGCCGAGTGTGTTGGTTCGGAAGCCCTGCTGCTGCTGCTGCTGCTGCTG 240
248 UGCUAGCCGAGUAGUGUGGUGCGGAAAGGCCUUGGUGUACUUGGUGUGUGUG 307
241 C 241
308 C 308

RESULT 9
US-08-397-220B-25
Sequence 25, Application US/08397220B
Patent No. 6284458

GENERAL INFORMATION:

APPLICANT: Anderson et al.

TITLE OF INVENTION: Compositions And Methods For Treatment
Of Hepatitis C Virus-Associated Diseases

NUMBER OF SEQUENCES: 98

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jane Massey Licata, Esq.

STREET: 210 Lake Drive East, Suite 201

CITY: Cherry Hill

STATE: NJ

COUNTRY: USA

ZIP: 08002

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE

COMPUTER: IBM 486

OPERATING SYSTEM: WINDOWS FOR WORKGROUPS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/397,220B

FILING DATE: 09-Mar-1995

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP93/01293

FILING DATE: 10-Sep-93
APPLICATION NUMBER: JP 5-87195
FILING DATE: 14-Apr-93
APPLICATION NUMBER: 07/945,289
FILING DATE: 10-Sep-92
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: ISPH-0031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 686
TYPE: nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: No
SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-08-397-220B-25

Query Match 100.0%; Score 241; DB 4; Length 686;
Best Local Similarity 79.3%; Pred. No. 1.9e-68;
Matches 191; Conservative 50; Mismatches 0; Indels 0; Gaps 0;

1 GCAGAAAGCGTGTAGCCATGCGGTAGTATGATGCTGCGAGCCCTCCAGAGACCCCT 60
68 GCAGAAAGCGTGTAGCCATGCGGTAGTATGATGCTGCGAGCCCTCCAGAGACCCCT 127
61 CCCGGAGAGCCATAGTGTGCGGAACCGGTGAGTACACCGGAATTGCCAGAGAC 120
128 CCCGGAGAGCCATAGTGTGCGGAACCGGTGAGTACACCGGAATTGCCAGAGAC 187
121 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 180
188 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 247
181 TGCTAGCCGAGTGTGTTGGTTCGGAAGCCCTGCTGCTGCTGCTGCTGCTGCTG 240
248 UGCUAGCCGAGUAGUGUGGUGCGGAAAGGCCUUGGUGUACUUGGUGUGUGUG 307
241 C 241
308 C 308

RESULT 10

US-08-650-093C-25
Sequence 25, Application US/08650093C
Patent No. 6381542

GENERAL INFORMATION:

APPLICANT: Kevin P. Anderson et al.

TITLE OF INVENTION: Compositions And Methods For Treatment Of
Hepatitis C Virus-Associated Diseases

NUMBER OF SEQUENCES: 118

CORRESPONDENCE ADDRESS:

ADDRESSEE: LICATA & TYRRELL P.C.

STREET: 66 E. Main Street

CITY: Marlton

STATE: NJ

COUNTRY: USA

ZIP: 08053

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: WORDPERFECT 6.1 for Windows

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/650,093C

FILING DATE: 17-May-1996

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

Best Local Match	100.0%;	Score 241;	DB 4;	Length 686;
Best Local Similarity	79.38;	Pred. No. 1.9e-68;		
Matches 191;	Conservative 50;	Mismatches 0;		

RESULT 11
US-08-474-700B-45
Sequence 45, Application US/08474700B
Patent No. 7,601,000

GENERAL INFORMATION:
APPLICANT: Wands, Jack
APPLICANT: Wakita, Takaji
APPLICANT: Moradpour, Darius
TITLE OF INVENTION: ANTISENSE INHIBITION
TITLE OF INVENTION: VIRUS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
OPERATING SYSTEM: IBM PS/2 Model 50Z or 55SX
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,700B

Very match	100.0%	Score 241;	DB 3;	Length 760;
Best Local Similarity	100.0%	Pred. No. 2e-68;		
Matches 241; Conservative	0;	Mismatches	0;	Total

RESULT: 12
US-08-869-380-1
; Sequence 1, Application US/08869380

APPLICANT: Pachuk, Catherine J.
 APPLICANT: Mandis, Jack
 APPLICANT: Wakita, Takaji
 APPLICANT: Zurawski, Jr., Vincent R.
 APPLICANT: Coney, Leslie R.
 TITLE OF INVENTION: Hepatitis C Virus Vaccine
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
 ADDRESS: No. 6235888r1s
 STREET: One Liberty Place, 46th floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/869,380
 FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/318,248
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: APOL-0189
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 923 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 342..914
US-08-869-380-1

Query Match 100.0%; Score 241; DB 4; Length 923;
Best Local Similarity 100.0%; Pred. No. 2,1e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGCGTTAGTATGATGTCGTGACGCTCCAGAGACCCCCCT 60
DB 68 GCAGAAAGCGTCTAGCCATGCGCGTTAGTATGATGTCGTGACGCTCCAGAGACCCCCCT 127

QY 61 CCCGGAGAGCCATAGTGTGTCGGGAACCGGTAGTACCGGATGCGAGAGAGACC 120
DB 128 CCCGGAGAGCCATAGTGTGTCGGGAACCGGTAGTACCGGATGCGAGAGAGACC 187

QY 121 GGGTCTTTCTTGGATCAACCCGCTCATGCTGAGATTGGGCGTCCCGCGAGAC 180
DB 188 GGGTCTTTCTTGGATCAACCCGCTCATGCTGAGATTGGGCGTCCCGCGAGAC 247

QY 181 TGTAGCCGAGTAGTGTGGTCCGGAAGCGCTTGTGTAAGTGTGCTGATAGGCTGTTG 240
DB 248 TGTAGCCGAGTAGTGTGGTCCGGAAGCGCTTGTGTAAGTGTGCTGATAGGCTGTTG 307

QY 241 C 241
DB 308 C 308

RESULT 13
PCT-US95-13552-14
Sequence 14, Application PC/TUS9513552
GENERAL INFORMATION:
APPLICANT: Wands, Jack
APPLICANT: Tokushige, Katsutoshi
APPLICANT: Makita, Takaji
APPLICANT: Pachuk, Catherine J.
APPLICANT: Zurawski, Jr., Vincent R.
TITLE OF INVENTION: HEPATITIS VIRUS VACCINES
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13552
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/318,248
FILING DATE: 05-OCT-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/467,859
FILING DATE: 06-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: APOL-0238
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 923 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 342..914
PCT-US95-13552-14

Query Match 100.0%; Score 241; DB 5; Length 923;
Best Local Similarity 100.0%; Pred. No. 2,1e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGCGTTAGTATGATGTCGTGACGCTCCAGAGACCCCCCT 60
DB 68 GCAGAAAGCGTCTAGCCATGCGCGTTAGTATGATGTCGTGACGCTCCAGAGACCCCCCT 127

QY 61 CCCGGAGAGCCATAGTGTGTCGGGAACCGGTAGTACCGGATGCGAGAGAGACC 120
DB 128 CCCGGAGAGCCATAGTGTGTCGGGAACCGGTAGTACCGGATGCGAGAGAGACC 187

QY 121 GGGTCTTTCTTGGATCAACCCGCTCATGCTGAGATTGGGCGTCCCGCGAGAC 180
DB 188 GGGTCTTTCTTGGATCAACCCGCTCATGCTGAGATTGGGCGTCCCGCGAGAC 247

QY 181 TGTAGCCGAGTAGTGTGGTCCGGAAGCGCTTGTGTAAGTGTGCTGATAGGCTGTTG 240
DB 248 TGTAGCCGAGTAGTGTGGTCCGGAAGCGCTTGTGTAAGTGTGCTGATAGGCTGTTG 307

QY 241 C 241
DB 308 C 308

RESULT 14
US-08-324-977-3
Sequence 3, Application US/08324977
Patent No. 5747339
GENERAL INFORMATION:
APPLICANT: OKAYAMA, HIROTO
APPLICANT: FUKE, ISAO
APPLICANT: MORI, CHISATO
APPLICANT: TAKAMIZAWA, AKAHISA
APPLICANT: YOSHIDA, IWAO
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Westerman, Hattori, McIreland &
STREET: 1725 K St. N.W. Suite 1000

CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/324,977
FILING DATE: 18-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/099,706
FILING DATE: 30-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/769,996
FILING DATE: 02-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/635,451
FILING DATE: 28-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Stevens-Smith, Theresa M.
REGISTRATION NUMBER: 36,281
REFERENCE/DOCKET NUMBER: 900703D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1499 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA from genomic RNA
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1499
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..1499
OTHER INFORMATION: /note= "sequence = 1 - 1499 of SEQ
US-08-324-977-3
Query Match 100.0%; Score 241; DB 1; Length 1499;
Best Local Similarity 100.0%; Pred. No. 2,4e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCATGCGCTTGTATGAGTGTGCGACGCTCCAGAGACCCCTT 60
DB 59 GCAGAAAGCGCTAGCATGCGCTTGTATGAGTGTGCGACGCTCCAGAGACCCCTT 118
QY 61 CCCGGAGAGACCATAGTGTGCGGAGACCGGTGAGTACACCGGAATTGCCAGAGAC 120
DB 119 CCCGGAGAGACCATAGTGTGCGGAGACCGGTGAGTACACCGGAATTGCCAGAGAC 178
QY 121 GGGTCTTCTTGATGATCAACCCGCTCATGCTGAGATTTGGCGTCCCGCGAGAC 180
DB 179 GGGTCTTCTTGATGATCAACCCGCTCATGCTGAGATTTGGCGTCCCGCGAGAC 238
QY 181 TCGTAGCCAGTAGTGTGGTGGCGGAAGGCTTGTGTACTGCTATAGGTGCTTG 240

DB 239 TCGTAGCCAGTAGTGTGGTGGCGGAAGGCTTGTGTACTGCTATAGGTGCTTG 298
QY 241 C 241
DB 299 C 299
RESULT 15
US-08-384-616-3
Sequence 3, Application US/08384616
Patent No. 58477101
GENERAL INFORMATION:
APPLICANT: OKAYAMA, Hiroto
APPLICANT: FUKE, Isao
APPLICANT: MORI, Chisato
APPLICANT: TAKAMIZAWA, Akahisa
APPLICANT: YOSHIDA, Itao
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
TITLE OF INVENTION: cDNA AND ANTIGEN POLYPEPTIDE
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Westernman, Hattori, Mclelland &
STREET: 1725 K St. N.W. Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,616
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/769,996
FILING DATE: 02-OCT-1991
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/635,451
FILING DATE: 28-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Stevens-Smith, Theresa M.
REGISTRATION NUMBER: 36,281
REFERENCE/DOCKET NUMBER: 900703B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1499 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA from genomic RNA
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1499
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..1499
OTHER INFORMATION: /note= "sequence = 1 - 1499 of SEQ

OTHER INFORMATION: ID: NO: 1"
US-08-384-616-3

Query Match 100.0%; Score 241; DB 2; Length 1499;
Best Local Similarity 100.0%; Pred. No. 2.4e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 GCAGAAAGCGCTAGCCATGCGCTTATGATGTCGTGACGCTCCAGGACCCCT 60
   |||||||
Db 59 GCAGAAAGCGCTAGCCATGCGCTTATGATGTCGTGACGCTCCAGGACCCCT 118
   |||||||
OY 61 CCCGGAGAGCCATAGTGTCTGCGAACCGGTAGTACACCGGATTCAGAGACGACC 120
   |||||||
Db 119 CCCGGAGAGCCATAGTGTCTGCGAACCGGTAGTACACCGGATTCAGAGACGACC 178
   |||||||
OY 121 GGGTCCTTCTTGATCAACCCGCTCAATGCTGAGATTGGCGCTGCCCGGAGAC 180
   |||||||
Db 179 GGGTCCTTCTTGATCAACCCGCTCAATGCTGAGATTGGCGCTGCCCGGAGAC 238
   |||||||
OY 181 TGCTAGCCGAGTGTGTTGGTCCGAAAGCCCTTGTGTAAGTGGTGTG 240
   |||||||
Db 239 TGCTAGCCGAGTGTGTTGGTCCGAAAGCCCTTGTGTAAGTGGTGTG 298
   |||||||
OY 241 C 241
   |
Db 299 C 299
```

Search completed: July 10, 2003, 20:28:04
Job time : 109.773 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:13:45 ; Search time 593.333 Seconds
(without alignments)
900.761 Million cell updates/sec

Title: US-10-087-631B-3

Perfect score: 33

Sequence: 1 cgggtactacacccgtccgcagaccactatgac 33

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues 32308132

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database :

EST:
1: em_estb:
2: em_esth:
3: em_esti:
4: em_estm:
5: em_estov:
6: em_estpl:
7: em_estro:
8: em_hic:
9: gb_estl:
10: gb_est2:
11: gb_hic:
12: gb_est3:
13: gb_est4:
14: gb_est5:
15: em_estfun:
16: em_estom:
17: gb_gss:
18: em_gss_hum:
19: em_gss_inv:
20: em_gss_pln:
21: em_gss_vrt:
22: em_gss_fun:
23: em_gss_mam:
24: em_gss_mus:
25: em_gss_other:
26: em_gss_pro:
27: em_gss_rtd:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.2	61.2	370	10	AV398524
2	20.2	61.2	466	9	AU004535
3	20.2	61.2	616	10	AM495169
4	20.2	61.2	650	10	AV398282
5	20.2	61.2	667	10	AV398654
6	20.2	61.2	671	10	AV398957

7	20.2	61.2	675	10	AV398322	AV398322
8	20.2	61.2	677	10	AV398498	AV398498
9	20.2	61.2	677	10	AV399052	AV399052
10	20.2	61.2	679	10	AV398803	AV398803
11	20.2	61.2	693	10	AV398253	AV398253
12	20.2	61.2	701	10	AV398450	AV398450
13	20.2	61.2	713	10	AV401743	AV401743
14	20.2	61.2	735	10	AV398164	AV398164
15	20.2	61.2	790	10	AV398482	AV398482
16	20.2	61.2	1845	11	AK005732	AK005732
17	20	60.6	467	10	AM195726	AM195726
18	20	60.6	772	12	BF203754	BF203754
19	20	60.6	1675	12	BG025626	BG025626
20	19.8	60.0	214	10	AV300075	AV300075
21	19.8	60.0	421	9	AL835471	AL835471
22	19.8	60.0	482	14	B0552508	B0552508
23	19.8	60.0	507	9	AL834604	AL834604
24	19.8	60.0	567	17	A0399676	A0399676
25	19.8	60.0	589	14	C78921	C78921
26	19.8	60.0	744	10	BE260962	BE260962
27	19.8	60.0	1033	11	AK018050	AK018050
28	19.6	59.4	395	17	A2244546	A2244546
29	19.6	59.4	644	12	BG333148	BG333148
30	19.6	59.4	792	12	BF218319	BF218319
31	19.6	59.4	1134	12	BG254464	BG254464
32	19.4	58.8	213	9	BB165683	BB165683
33	19.4	58.8	373	9	AA953786	AA953786
34	19.4	58.8	576	9	AU148437	AU148437
35	19.4	58.8	581	9	AU171278	AU171278
36	19.4	58.8	594	13	B0082203	B0082203
37	19.4	58.8	646	13	B0111227	B0111227
38	19.4	58.8	666	9	AU178284	AU178284
39	19.4	58.8	668	13	BJ487077	BJ487077
40	19.4	58.8	670	13	BJ007208	BJ007208
41	19.4	58.8	675	9	AU169949	AU169949
42	19.4	58.8	680	13	B013902	B013902
43	19.4	58.8	697	13	B020967	B020967
44	19.4	58.8	734	12	BG503487	BG503487
45	19.4	58.8	755	9	AU169952	AU169952

ALIGNMENTS

RESULT 1
LOCUS AV398524 370 bp mRNA linear EST 05-FEB-2000
DEFINITION AV398524 Bombyx mori ovary BmPV infected; 2 hr after inoculation
Bombyx mori CDNA clone NV021911X T3, mRNA sequence.

ACCESSION AV398524
VERSION AV398524.1 GI:6902176
KEYWORDS EST
SOURCE domestic silkworm.
ORGANISM Bombyx mori

REFERENCE
AUTHORS Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.
TITLE Bombyx mori CDNA
JOURNAL Unpublished (2000)
COMMENT Contact: Mita K

Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@nirs.go.jp
method: uni-directional, sequence direction: sequenced from T3 primer
(5' -> 3')

Project: 'Silkworm Genome Program in MAF, and Research for the
Future Program in JSPS', see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers
1. 370

FEATURES
source

```

/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone_id="Bombyx mori ovary BmNPV infected; 2 hr after
inoculation"
/tissue_type="ovary"
/cell_type="BmNPV infected cell"
/dev_stage="BmNPV infected; 2 hr after inoculation"
BASE COUNT
ORIGIN
63 a 102 c 115 g 89 t 1 others

Query Match
Best Local Similarity 61.2%; Score 20.2; DB 10; Length 370;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 CGGTGATCAGCTCCGTCGAGACCACTATGCG 33
82 CGCGGAACTCTCCGTCCTCGGCGCCCACTATCGC 114

RESULT 2
AU004535 466 bp mRNA linear EST 19-JAN-1999
LOCUS Bombyx mori p50(Dalzo) Bombyx mori CDNA clone ws20577,
DEFINITION mRNA sequence.
ACCESSION AU004535.1 GI:4161906
VERSION AU004535.1
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 466)
Mita,K., Moriyama,M., Shimada,T., Okano,K. and Maeda,S.
Establishment of CDNA database of Bombyx mori
Unpublished (1999)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmitsuen@irs.go.jp
PROJECT = 'CREST project by JST'
FEATURES
source
1. 466
/organism="Bombyx mori"
/strain="p50(Dalzo)"
/db_xref="taxon:7091"
/clone_id="ws20577"
/clone_id="Bombyx mori p50(Dalzo)"
BASE COUNT
ORIGIN
82 a 125 c 147 g 112 t

Query Match
Best Local Similarity 61.2%; Score 20.2; DB 9; Length 466;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 CGGTGATCAGCTCCGTCGAGACCACTATGCG 33
91 CGCGGAACTCTCCGTCCTCGGCGCCCACTATCGC 123

RESULT 3
AM495169 616 bp mRNA linear EST 24-FEB-2000
LOCUS Bombyx mori p50(Dalzo) Bombyx mori CDNA clone
DEFINITION U1-M-BH3-aur-h-07-0-UI-s1 NIH-BMAP_M_S4 Mus musculus CDNA clone
ACCESSION AM495169
VERSION AM495169.1 GI:7065415
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 616)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestrail.nih.gov
Oligo-dt track not found, Not I site shown in beginning of sequence
is likely internal to the message. CDNA library preparation: M.B.
Soares lab clone distribution: Researchers may obtain BMAP CDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP CDNA whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA-NO.

FEATURES
SOURCE

Location/Qualifiers
1. 616
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U1-M-BH3-aur-h-07-0-UI"
/clone_id="NIH-BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The
NIH-BMAP_M_S4 library is a subtracted library of a series
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH-BMAP_M_S3, NIH-BMAP_M_S3.1,
NIH-BMAP_M_S3.2, NIH-BMAP_M_S3.3, NIH-BMAP_M_S3.4,
NIH-BMAP_M_S4. The subtracted library
(NIH-BMAP_M_S4) was constructed as follows: PCR amplified
CDNA inserts from NIH-BMAP_M_S3.3, NIH-BMAP_M_S3.2, and
NIH-BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH-BMAP_M_S3.3, NIH-BMAP_M_S3.2, and NIH-BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH-BMAP_M_S4 library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=NIH-BMAP_M_S4
TAG_TISSUE=Basal ganglia
TAG_SEQ=GTATAC

BASE COUNT
ORIGIN
137 a 179 c 148 g 152 t

```

Query Match
Best Local Similarity 61.2%; Score 20.2; DB 10; Length 616;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 CGGTGATCAGCTCCGTCGAGACCACTATGCG 33
165 CAGGTATCAGAGCTCCGTCGAGACCACTATGCG 197

```

[illegible][illegible]

RESULT 7
AV398322 675 bp mRNA linear EST 05-FEB-2000
LOCUS AV398322 Bombyx mori ovary BmPV infected; 2 hr after inoculation
DEFINITION Bombyx mori cDNA clone NV021679 T3, mRNA sequence.
ACCESSION AV398322.1 GI:6901974
VERSION EST.
KEYWORDS domestic silkworm.
SOURCE Bombyx mori
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 675)
AUTHORS Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
TITLE Bombyx mori cDNA
JOURNAL Unpublished (2000)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmits@nirs.go.jp
method:uni-directional, sequence direction:sequenced from T3 primer

BASE COUNT 122 a 190 c 206 g 157 t

ORIGIN

Query Match 61.2%; Score 20.2; DB 10; Length 675;
Best Local Similarity 75.8%; Pred. No. 6.3e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CGGTGCTACTACCGTCCGCGACGACCTATGCG 33
82 CGCGAAGCTCTCCGTCGTCGCGCCGACCTATGCG 114

Db

RESULT 8
AV398498 677 bp mRNA linear EST 05-FEB-2000
LOCUS AV398498 Bombyx mori ovary BmPV infected; 2 hr after inoculation
DEFINITION Bombyx mori cDNA clone NV021881 T3, mRNA sequence.
ACCESSION AV398498
VERSION AV398498.1 GI:6902150
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 677)
AUTHORS Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
TITLE Bombyx mori cDNA
JOURNAL Unpublished (2000)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmits@nirs.go.jp
method:uni-directional, sequence direction:sequenced from T3 primer

BASE COUNT 121 a 189 c 205 g 157 t

ORIGIN

Query Match 61.2%; Score 20.2; DB 10; Length 677;
Best Local Similarity 75.8%; Pred. No. 6.3e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CGGTGCTACTACCGTCCGCGACGACCTATGCG 33
82 CGCGAAGCTCTCCGTCGTCGCGCCGACCTATGCG 114

Db

FEATURES
source

1.675
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV021679"
/clone_lib="Bombyx mori ovary BmPV infected; 2 hr after inoculation"
/tissue_type="ovary"
/cell_type="BmPV cultured cell"
/dev_stage="BmPV infected; 2 hr after inoculation"

REFERENCE
AUTHORS Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
TITLE Bombyx mori cDNA
JOURNAL Unpublished (2000)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmits@nirs.go.jp
method:uni-directional, sequence direction:sequenced from T3 primer

Project="Silkworm Genome Program in MARF, and Research for the Future Program in JSPS", see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers

(5' -> 3')
Project="Silkworm Genome Program in MARF, and Research for the Future Program in JSPS", see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers

1.677
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV021881"
/clone_lib="Bombyx mori ovary BmPV infected; 2 hr after inoculation"
/tissue_type="ovary"
/cell_type="BmPV cultured cell"
/dev_stage="BmPV infected; 2 hr after inoculation"

BASE COUNT 125 a 191 c 203 g 158 t

ORIGIN

Query Match 61.2%; Score 20.2; DB 10; Length 677;
Best Local Similarity 75.8%; Pred. No. 6.3e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CGGTGCTACTACCGTCCGCGACGACCTATGCG 33
77 CGCGAAGCTCTCCGTCGTCGCGCCGACCTATGCG 109

Db

RESULT 9
AV399052 677 bp mRNA linear EST 05-FEB-2000
LOCUS AV399052 Bombyx mori ovary BmPV infected; 6 hr after inoculation
DEFINITION Bombyx mori cDNA clone NV060603 T3, mRNA sequence.
ACCESSION AV399052
VERSION AV399052.1 GI:6902704
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 677)
AUTHORS Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
TITLE Bombyx mori cDNA
JOURNAL Unpublished (2000)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmits@nirs.go.jp
method:uni-directional, sequence direction:sequenced from T3 primer
(5' -> 3')
Project="Silkworm Genome Program in MARF, and Research for the Future Program in JSPS", see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers

1.677
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV060603"
/clone_lib="Bombyx mori ovary BmPV infected; 6 hr after inoculation"
/tissue_type="ovary"
/cell_type="BmPV cultured cell"
/dev_stage="BmPV infected; 6 hr after inoculation"

BASE COUNT 121 a 189 c 205 g 157 t

ORIGIN

Query Match 61.2%; Score 20.2; DB 10; Length 677;
Best Local Similarity 75.8%; Pred. No. 6.3e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CGGTGCTACTACCGTCCGCGACGACCTATGCG 33
82 CGCGAAGCTCTCCGTCGTCGCGCCGACCTATGCG 114

Db

DB 82 CGCGAAGCTCCGTCGCGGCCACTATGCG 114

RESULT 13
AV401743 713 bp mRNA linear EST 05-FEB-2000
LOCUS AV401743
DEFINITION Clone hes00104 T3, mRNA sequence.
ACCESSION AV401743
VERSION AV401743.1 GI:6905395
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 713)
AUTHORS Mita,K., Moriyama,M., Shimada,T., Okano,K. and Maeda,S.
TITLE Bombyx mori cDNA
JOURNAL Unpublished (2000)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

BASE COUNT
ORIGIN 134 a 201 c 211 g 163 t 4 others

FEATURES
Source
1. 713
/organism="Bombyx mori"
/strain="C108"
/db_xref="taxon:7091"
/clone="hes00104"
/clone_lib="Bombyx mori C108 spinning stage day-0"
/sex="female/male mixed"
/cell_type="hemocyte"
/dev_stage="spinning stage day-0"

Query Match 61.2% Score 20.2; DB 10; Length 713;
Best Local Similarity 75.8%; Pred. No. 6.4e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 57 CGCGAAGCTCCGTCGCGGCCACTATGCG 89

RESULT 14
AV398164 735 bp mRNA linear EST 05-FEB-2000
LOCUS AV398164
DEFINITION Bombyx mori ovary BmNPV infected; 2 hr after inoculation
ACCESSION AV398164
VERSION AV398164.1 GI:6901816
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 735)
AUTHORS Mita,K., Moriyama,M., Shimada,T., Okano,K. and Maeda,S.
TITLE Bombyx mori cDNA
JOURNAL Unpublished (2000)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

BASE COUNT
ORIGIN 150 a 224 c 235 g 181 t

FEATURES
Source
1. 790
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV021859"
/clone_lib="Bombyx mori ovary BmNPV infected; 2 hr after inoculation"
/tissue_type="ovary"
/cell_type="BmNPV infected cell"
/dev_stage="BmNPV infected; 2 hr after inoculation"

Query Match 61.2% Score 20.2; DB 10; Length 790;
Best Local Similarity 75.8%; Pred. No. 6.6e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 1 CGGTGACTACCGTCGCGAGACACTATGCG 33

RESULT 15
AV398482 790 bp mRNA linear EST 05-FEB-2000
LOCUS AV398482
DEFINITION Bombyx mori ovary BmNPV infected; 2 hr after inoculation
ACCESSION AV398482
VERSION AV398482.1 GI:6902134
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 790)
AUTHORS Mita,K., Moriyama,M., Shimada,T., Okano,K. and Maeda,S.
TITLE Bombyx mori cDNA
JOURNAL Unpublished (2000)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@nirs.go.jp
method:uni-directional, sequence direction:sequenced from T3 primer
(5' -> 3')
Project="Silkworm Genome Program in MAFF, and Research for the
Future Program in JSPS", see 'Silkbase',
Location/Qualifiers
Location/Qualifiers
1. 790
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV021859"
/clone_lib="Bombyx mori ovary BmNPV infected; 2 hr after inoculation"
/tissue_type="ovary"
/cell_type="BmNPV infected cell"
/dev_stage="BmNPV infected; 2 hr after inoculation"

Query Match 61.2% Score 20.2; DB 10; Length 790;
Best Local Similarity 75.8%; Pred. No. 6.6e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 72 CGCGAAGCTCCGTCGCGGCCACTATGCG 104

RESULT 16
AV398482 790 bp mRNA linear EST 05-FEB-2000
LOCUS AV398482
DEFINITION Bombyx mori ovary BmNPV infected; 2 hr after inoculation
ACCESSION AV398482
VERSION AV398482.1 GI:6902134
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 790)
AUTHORS Mita,K., Moriyama,M., Shimada,T., Okano,K. and Maeda,S.
TITLE Bombyx mori cDNA
JOURNAL Unpublished (2000)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@nirs.go.jp
method:uni-directional, sequence direction:sequenced from T3 primer
(5' -> 3')
Project="Silkworm Genome Program in MAFF, and Research for the
Future Program in JSPS", see 'Silkbase',
Location/Qualifiers
Location/Qualifiers
1. 790
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV021859"
/clone_lib="Bombyx mori ovary BmNPV infected; 2 hr after inoculation"
/tissue_type="ovary"
/cell_type="BmNPV infected cell"
/dev_stage="BmNPV infected; 2 hr after inoculation"

Query Match 61.2% Score 20.2; DB 10; Length 790;
Best Local Similarity 75.8%; Pred. No. 6.6e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 1 CGGTGACTACCGTCGCGAGACACTATGCG 33

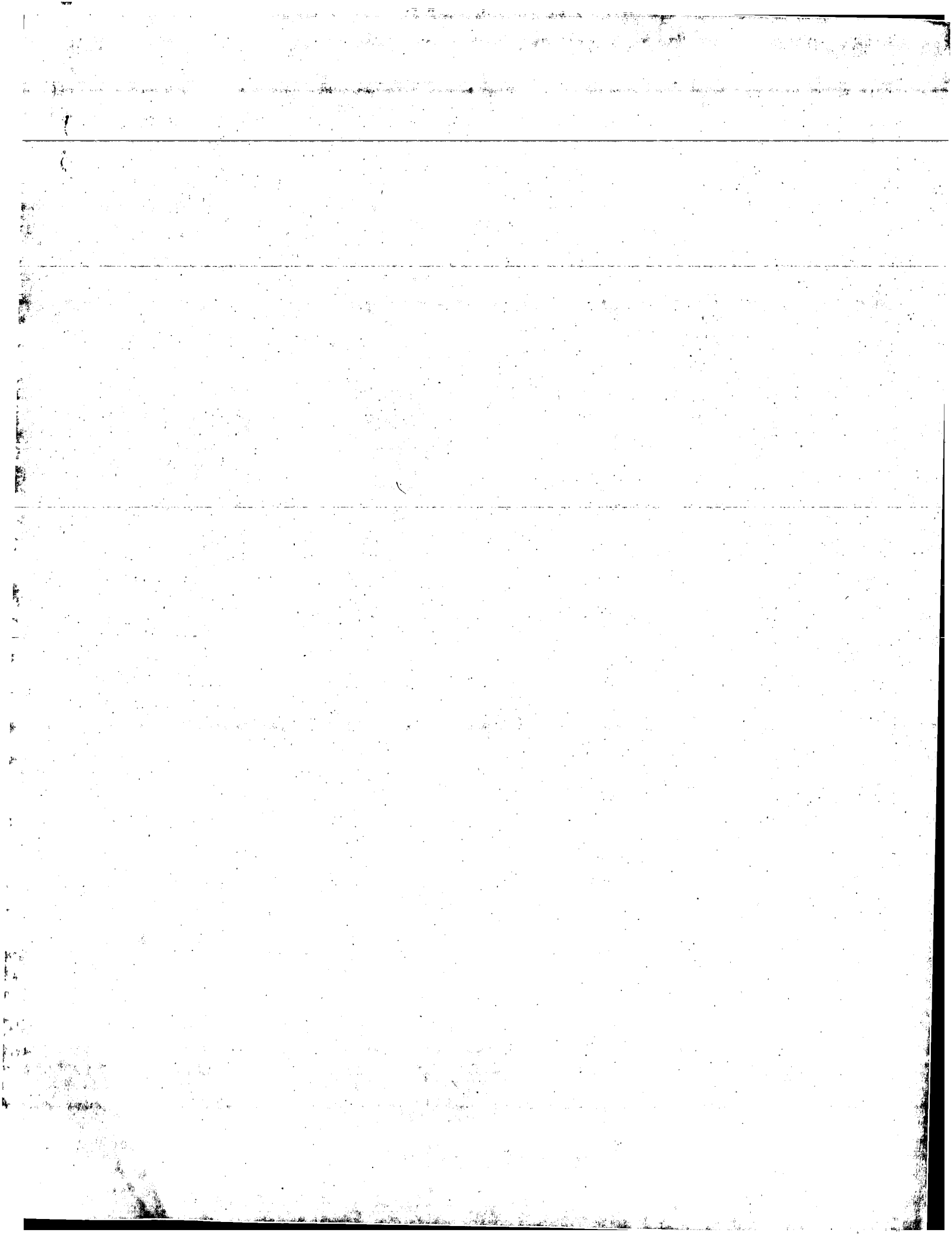
Fri Jul 11 15:29:34 2003

us-10-087-631b-3.rst

Page 7

Db 66 CGCGAAGCTCTCCGTTCCGTCGCGCCCACTATCGC 98

Search completed: July 11, 2003, 02:25:28
Job time : 600.333 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:13:45 ; Search time 539.393 Seconds
(Without alignments)
900.761 Million cell updates/sec

Title: US-10-087-631b-4

Perfect score: 30
Sequence: 1 tggactacgcttcgtgcacacaccccttc 30

Scoring table: IDENTITY-NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlm:*
5: em_estloy:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_hum:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	21	70.0	354	13	BI014511	PMO-ET020
2	21	70.0	912	21	BE871492	601450002
3	21	70.0	1008	13	BM468259	ACGENCOURT
4	20.4	68.0	280	9	AV047458	AV047458
5	20.4	68.0	296	9	AA432035	AA432035
6	20.4	68.0	361	9	AA758968	AA758968

Result	Score	Match	Length	DB	ID	Description
C 7	20.4	68.0	385	9	AA396639	AA396639
8	20.4	68.0	509	9	AA398952	AA398952
C 9	20.4	68.0	539	9	AI601049	AI601049
10	20.4	68.0	542	14	BQ328104	BQ328104
11	20.4	68.0	582	12	BG383655	BG383655
C 12	20.4	68.0	705	9	AV244148	AV244148
13	20.4	68.0	800	17	CNS03PGU	CNS03PGU
14	20.4	68.0	1100	17	CNS05LFI	CNS05LFI
15	20.4	68.0	1188	12	BG250312	BG250312
C 16	20.4	68.0	1426	12	BG848013	BG848013
C 17	20.2	67.3	118	17	AZ387734	AZ387734
C 18	20.2	67.3	336	14	W98110	W98110
C 19	20.2	66.7	516	14	BO602301	BO602301
C 20	20.2	66.7	571	12	BF099337	BF099337
C 21	20.2	66.7	585	12	BF427489	BF427489
C 22	20.2	66.7	635	10	BB655806	BB655806
23	20.2	66.7	954	17	BH728734	BH728734
C 24	20.2	66.7	966	17	CNS02E61	CNS02E61
C 25	20.2	66.7	981	17	CNS06206	CNS06206
C 26	19.8	66.0	909	12	BG399084	BG399084
C 27	19.6	65.3	302	10	BB098595	BB098595
C 28	19.6	65.3	447	17	AQ368408	AQ368408
C 29	19.6	65.3	525	17	AQ368431	AQ368431
C 30	19.6	65.3	554	17	AZ451858	AZ451858
C 31	19.6	65.3	584	14	C79123	C79123
C 32	19.6	65.3	663	17	AG075408	AG075408
C 33	19.6	65.3	720	17	AG165031	AG165031
C 34	19.6	65.3	902	14	BO430998	BO430998
C 35	19.6	65.3	936	17	BM472701	BM472701
C 36	19.6	65.3	1034	13	CNS04TV0	CNS04TV0
C 37	19.4	64.7	275	10	BE047681	BE047681
C 38	19.4	64.7	327	10	BB222109	BB222109
C 39	19.4	64.7	386	17	AZ892007	AZ892007
C 40	19.4	64.7	527	12	BG154115	BG154115
C 41	19.4	64.7	533	10	BE681567	BE681567
C 42	19.4	64.7	541	17	BH687462	BH687462
C 43	19.4	64.7	550	14	BO339788	BO339788
C 44	19.4	64.7	560	9	AI341663	AI341663
C 45	19.4	64.7	584	12	BF844863	BF844863

ALIGNMENTS

RESULT 1
BI014511/c
LOCUS
DEFINITION
PMO-ET0208-030201-005-C04 ET0208 Homo sapiens CDNA, mRNA sequence.
BI014511
ACCESSION
BI014511.1 GI:14418582
VERSION
BI014511.1
KEYWORDS
EST.
SOURCE
human
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 354)
AUTHORS
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Negai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM0&t2=PM0-ET0208-030201-005-c04&t3=2001-02-03&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 27
 High quality sequence stop: 354
 Location/Qualifiers
 1..354

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="ET0208"
 /dev_stage="Adult"
 /note="Organ: lung, tumor; Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESNES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 116 a 53 c 115 g 70 t

BASE COUNT
 ORIGIN
 Query Match 70.0%; Score 21; DB 13; Length 354;
 Best Local Similarity 82.8%; Pred. No. 3.2e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 2 GGACTGAGTCTGTGTCATCTCACCCTCT 30
 11 ||||| ||||| ||||| ||||| |||||
 Db 66 GGCTCAGTACTGTGTCATCTCTCTCTCT 38

RESULT 2
 BE871492 912 bp mRNA linear EST 20-OCT-2000
 LOCUS 601450002F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3853777 5'
 DEFINITION BE871492
 mRNA sequence.
 BE871492.1 GI:10320268
 EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 912)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM9578 row: d column: 02
 High quality sequence start: 18
 High quality sequence stop: 551.
 Location/Qualifiers
 1..912

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NIH_MGC_65"
 /tissue_type="adrenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pCMV-SPO6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: oligo dT.
 Average insert size 1.8 kb. Library constructed by Life Technologies."
 Technology.

BASE COUNT 241 a 317 c 175 g 179 t
 ORIGIN
 Query Match 70.0%; Score 21; DB 12; Length 912;
 Best Local Similarity 82.8%; Pred. No. 4.8e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 2 GGACTGAGTCTGTGTCATCTCACCCTCT 30
 11 ||||| ||||| ||||| ||||| |||||
 Db 258 GGACTGAGTCTGTGTCATCTCACCCTCT 286

RESULT 3
 BM468259 1008 bp mRNA linear EST 05-FEB-2002
 LOCUS AGENCOURT 6475704 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5578052
 DEFINITION 5' mRNA sequence.
 BM468259
 ACCESSION BM468259.1 GI:18517301
 VERSION EST.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1008)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: ATCC/DCID/DTF
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM12332 row: h column: 21
 High quality sequence stop: 707.
 Location/Qualifiers
 1..1008

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NIH_MGC_72"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SPO6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: oligo dT.
 Average insert size 2 kb. Library constructed by Life Technologies."
 Technology.

BASE COUNT 291 a 269 c 221 g 219 t 8 others
 ORIGIN
 Query Match 70.0%; Score 21; DB 13; Length 1008;
 Best Local Similarity 82.8%; Pred. No. 5e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 2 GGACTGAGTCTGTGTCATCTCACCCTCT 30
 11 ||||| ||||| ||||| ||||| |||||
 Db 109 GGACTGAGTCTGTGTCATCTCACCCTCT 137

RESULT 4
 AV047458 280 bp mRNA linear EST 23-NOV-1999
 LOCUS AV047458 Mus musculus adult C57BL/6J testis Mus musculus cDNA clone
 DEFINITION 1700067016, mRNA sequence.
 AV047458
 ACCESSION AV047458.2 GI:4867123
 VERSION EST.
 KEYWORDS house mouse.
 SOURCE
 ORGANISM Mus musculus

FEATURES

Insert Length: 1893 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 311.
Location/Qualifiers

BASE COUNT

75 a 98 c 99 g 89 t

Query Match 68.0%; Score 20.4; DB 9; Length 361;
Best Local Similarity 80.0%; Pred. No. 5.6e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 307 TGGCTCAGCTCCTTGGCTGATCTGACCTTCT 278
1 TGGCTCAGCTCCTTGGCTGATCTGACCTTCT 30

RESULT 7 385 bp mRNA linear EST 16-MAY-1997
AA399639 2193906.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:729946
LOCUS 3', mRNA sequence.
DEFINITION AA399639
ACCESSION AA399639
VERSION
KEYWORDS
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 385)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, R., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie,
T., Waterston, R. and Wilson, R.
Washu-Merck EST Project 1997
Unpublished (1997)

TITLE JOURNAL
JOURNAL
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu

FEATURES

This clone is available royalty-free through LIND; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 340.
Location/Qualifiers

1. 385
/organism="Homo sapiens"
/db_xref="GDB:5926475"
/db_xref="taxon:9606"
/clone="IMAGE:729946"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

BASE COUNT

74 a 106 c 107 g 98 t

Query Match 68.0%; Score 20.4; DB 9; Length 385;
Best Local Similarity 80.0%; Pred. No. 5.8e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 331 TGGCTCAGCTCCTTGGCTGATCTGACCTTCT 302
1 TGGCTCAGCTCCTTGGCTGATCTGACCTTCT 30

RESULT 8 509 bp mRNA linear EST 16-MAY-1997
AA398952 2193906.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:729946
LOCUS 5', mRNA sequence.
DEFINITION AA398952
ACCESSION AA398952
VERSION
KEYWORDS
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 509)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, R., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie,
T., Waterston, R. and Wilson, R.
Washu-Merck EST Project 1997
Unpublished (1997)

TITLE JOURNAL
JOURNAL
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu

This clone is available royalty-free through LIND; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 480.
Location/Qualifiers

1. 509
/organism="Homo sapiens"
/db_xref="GDB:5926475"
/db_xref="taxon:9606"
/clone="IMAGE:729946"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTCTTTTCTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 82 a 174 c 146 g 107 t
ORIGIN
Query Match 68.0%; Score 20.4; DB 9; Length 509;
Best Local Similarity 80.0%; Pred. No. 6.5e+02;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGGACTCAGCTCTGGTCATCTCACCCTTCT 30
 ||| ||||| ||||| ||| ||||| |||
 Db 253 TGGTCTCAGCTCTGGTCATCTCACCCTTGT 282

RESULT 9
 A1601049/c 539 bp mRNA linear EST 21-APR-1999
 LOCUS 486096E11.x1 486 - leaf primordia cDNA library from Hake lab zea
 DEFINITION mays cDNA, mRNA sequence.
 ACCESSION A1601049
 VERSION A1601049.1 GI:4610210
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 539)
 Maibot.V.
 Maize ESTs from various cDNA libraries sequenced at Stanford
 University
 Unpublished (1999)
 Contact: Maibot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: maibot@stanford.edu
 Plate: 486096 row: E column: 11.
 Location/Qualifiers
 1..539
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone_lib="486 - leaf primordia cDNA library from Hake
 lab"
 /tissue_type="leaf primordia"
 /dev_stage="P7-P11 leaf"
 /lab_host="E.coli XL1-Blue MFR"
 /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
 library"

BASE COUNT 166 a 92 c 146 g 134 t 1 others

ORIGIN

Query Match 68.0%; Score 20.4; DB 9; Length 539;
 Best Local Similarity 80.0%; Pred. No. 6.6e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGGACTCAGCTCTGGTCATCTCACCCTTCT 30
 ||||| ||| ||| ||||| ||||| |||||
 Db 340 TGGACACGCTACATCTCCTCCTCTCT 311

RESULT 10
 B0328104 542 bp mRNA linear EST 17-MAY-2002
 LOCUS MRL-RF0079-151200-003-B02 RT0079 Homo sapiens cDNA, mRNA sequence.
 DEFINITION B0328104
 ACCESSION B0328104.1 GI:20946200
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 542)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

TITLE
 JOURNAL
 MEDLINE
 COMMENT

CONTACT: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SF,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l-MRL12-MRL-RF0079-
 151200-003-b02<=2000-12-15<4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 14
 High quality sequence stop: 31.
 Location/Qualifiers
 1..542
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="RF0079"
 /dev_stage="Adult"
 /note="Organ: kidney, tumor; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 90 a 194 c 136 g 119 t 3 others

ORIGIN

Query Match 68.0%; Score 20.4; DB 14; Length 542;
 Best Local Similarity 80.0%; Pred. No. 6.6e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGGACTCAGCTCTGGTCATCTCACCCTTCT 30
 ||| ||||| ||||| ||| ||||| |||
 Db 479 TGGTCTCAGCTCTGGTCATCTCACCCTTGT 508

RESULT 11
 BG383655 582 bp mRNA linear EST 12-MAR-2001
 LOCUS 301851 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
 DEFINITION BG383655
 ACCESSION BG383655.1 GI:13308127
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 582)
 Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
 Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
 and Keeler, J.W.
 Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine
 Unpublished (2000)
 Contact: Smith TP
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt-trimmed with phred
 v0.980904.e. Vector identified by cross-match with the -minscore 18
 and -mismatch 12 options.

```

/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_1ib="MARC 1pIG"
/tissue_type="pooled"
/lab_host="DH10B"
/notes=Vector: PCMV SPORTS; Site_1: Notti; Site_2: Sali;
Library made from sorted tissue from day 11, 13, 15, 20
and 30 embryos.
205 c 121 g 176 t

```

68.0%	Score 20.4;	DB 12;	Length 582;
80.0%	Pred. No. 6.8e+02;		
ive	0;	Mismatches 6;	Indels 0;
			Gaps 0

QY	1	TGACATCAGTCCCTTGGTCATCTCACCCTTCT	30
Db	211	TGGTCTCTGGCCCTGGGCACCTCACCCTTCT	240

705 bp mRNA linear EST 31-AUG-2001
RIKEN full-length enriched, 0 day neonate head Mus
cDNA clone 14831428F09 3', mRNA sequence.

2 GI:15405234

Use:

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 to 705)

TITLE: RIKEN Mouse ESTs (Arakawa, T., et al. 2001, unpublished (2001))
JOURNAL: On Nov 4, 1999 this sequence version replaced
COMMENT:

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9422
Fax: 81-45-503-9216

Email: genome-resesgsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>,
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. *Genome Res.* 10 (10), 1617-1630 (2000)
Wegli, K., Fumieda, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. *Genome Res.*
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shingawa, A., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T., Ishii, Y. and Hayashizaki, Y.

Mapping of 19033 mouse cDNAs on mouse chromosomes. *J. Struct. Func. Genomics* 2 pre, L72-L86 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Location/Qualifiers

1. .705

source

FEATURES

FEATURES
source

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4831428F09"
/clone_1b="R1K1N full-length enriched, 0 day neonate"
head"

```

209 a	158 c	152 g	186 t
-------	-------	-------	-------

Query Match	68.0%;	Score 20.4;	DB 9;	Length 705;
Best Local Similarity	80.0%;	Pred. No. 7.4e+02;		
Matches	24;	Conservative	0;	Mismatches 6;
				Indels 0;
				Gaps 0

QY 1 TGGACTCAGTCCTTGGTCATCTCACCCTTCT 30
||| ||| | ||||| | ||||| ||
Db. 345 TGGTCTCCTTACTTGGTCTTCACACCTTCT 316

LOCUS	CN503PcU	800 bp	DNA	linear	GSS 17-MAY-2000
DEFINITION	Tetradodon nigroviridis genome survey sequence PPC-Or1 end of clone 045115 of library G from Tetradodon nigroviridis, genomic survey sequence.				
ACCESSION	AL284631				
VERSION	AL284631.1	GI:7975643			
KEYWORDS	GSS; genome survey sequence.				
SOURCE	Tetradodon nigroviridis.				
ORGANISM	Tetradodon nigroviridis				

1 (bases 1 to 800)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.

Fri Jul 11 15:29:35 2003

Search completed: July 11, 2003, 02:25:36
Job time : 547.393 secs

us-10-087-631b-4.rst

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 : Search time 14.7573 Seconds
(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631B-3

Perfect score: 33

Sequence: 1 cgggtactaccgcgttcgcagacactatggc 33

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_MA:*

- 1: /cgn2_6/p/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgn2_6/p/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgn2_6/p/ptodata/1/ina/6A.COMB.seq:*
- 4: /cgn2_6/p/ptodata/1/ina/6B.COMB.seq:*
- 5: /cgn2_6/p/ptodata/1/ina/PCTUS.COMB.seq:*
- 6: /cgn2_6/p/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	66.7	120	2	US-08-690-495-26
2	22	66.7	120	2	US-08-690-494-26
3	22	66.7	120	2	PCT-US95-07671-26
4	22	66.7	145	2	US-08-690-495-18
5	22	66.7	145	2	US-08-690-494-18
6	22	66.7	145	5	PCT-US95-07671-18
7	22	66.7	177	2	US-08-256-568B-55
8	22	66.7	177	2	US-08-256-568B-56
9	22	66.7	177	2	US-08-256-568B-57
10	22	66.7	177	2	US-08-256-568B-58
11	22	66.7	177	2	US-08-256-568B-61
12	22	66.7	177	2	US-08-256-568B-62
13	22	66.7	177	2	US-08-256-568B-63
14	22	66.7	177	2	US-08-256-568B-64
15	22	66.7	177	2	US-08-256-568B-65
16	22	66.7	177	2	US-08-256-568B-66
17	22	66.7	177	2	US-08-256-568B-67
18	22	66.7	177	2	US-08-256-568B-68
19	22	66.7	177	2	US-08-256-568B-69
20	22	66.7	177	2	US-08-256-568B-70
21	22	66.7	177	2	US-08-256-568B-72
22	22	66.7	177	2	US-08-256-568B-73
23	22	66.7	177	2	US-08-256-568B-74
24	22	66.7	177	2	US-08-256-568B-75
25	22	66.7	177	2	US-08-256-568B-77
26	22	66.7	177	2	US-08-256-568B-78
27	22	66.7	177	2	US-08-256-568B-79

c 28	22	66.7	177	2	US-08-256-568B-80	Sequence 80, Appl
c 29	22	66.7	177	4	US-09-038-369B-55	Sequence 55, Appl
c 30	22	66.7	177	4	US-09-038-369B-56	Sequence 56, Appl
c 31	22	66.7	177	4	US-09-038-369B-57	Sequence 57, Appl
c 32	22	66.7	177	4	US-09-038-369B-58	Sequence 58, Appl
c 33	22	66.7	177	4	US-09-038-369B-61	Sequence 61, Appl
c 34	22	66.7	177	4	US-09-038-369B-62	Sequence 62, Appl
c 35	22	66.7	177	4	US-09-038-369B-63	Sequence 63, Appl
c 36	22	66.7	177	4	US-09-038-369B-64	Sequence 64, Appl
c 37	22	66.7	177	4	US-09-038-369B-65	Sequence 65, Appl
c 38	22	66.7	177	4	US-09-038-369B-66	Sequence 66, Appl
c 39	22	66.7	177	4	US-09-038-369B-67	Sequence 67, Appl
c 40	22	66.7	177	4	US-09-038-369B-68	Sequence 68, Appl
c 41	22	66.7	177	4	US-09-038-369B-69	Sequence 69, Appl
c 42	22	66.7	177	4	US-09-038-369B-70	Sequence 70, Appl
c 43	22	66.7	177	4	US-09-038-369B-72	Sequence 72, Appl
c 44	22	66.7	177	4	US-09-038-369B-73	Sequence 73, Appl
c 45	22	66.7	177	4	US-09-038-369B-74	Sequence 74, Appl

ALIGNMENTS

RESULT 1
US-08-690-495-26
Sequence 26, Application US/08690495
Patent No. 5876924

GENERAL INFORMATION:
APPLICANT: Zhang, David Y., Brandwein, Margaret
TITLE OF INVENTION: NUCLEIC ACID AMPLIFICATION METHOD:
TITLE OF INVENTION: HYBRIDIZATION SIGNAL AMPLIFICATION METHOD (HSAM)
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10112-0228

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Fasteo Version #1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/690,495
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MacLeod, Janet M.
REGISTRATION NUMBER: 35,263
REFERENCE/DOCKET NUMBER: 29545-A-PCT/USA-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-765-2519
TELEFAX: 212-765-2519
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 120 base pairs
TYPE: nucleic acid
TOPOLOGY: linear
STRANDEDNESS: single
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1..120
US-08-690-495-26

Query Match 66.7%; Score 22; DB 2; Length 120;
Best local similarity 97.1%; Pred. No. 0.27;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
CGGTGTACTCACCGTTCGCGAGACACTATGGC 33
|||||

Db 64 CGGTGTACTCACCAGTCCGACGACCATATGCG 97

RESULT 2

US-08-690-494-26
Sequence 26 Application US/08690494
Patent No. 5942391

GENERAL INFORMATION:

APPLICANT: Zhang, David Y., Brandwein, Margaret
TITLE OF INVENTION: NUCLEIC ACID AMPLIFICATION METHOD:
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
ZIP: 10112-0228COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Fasteq Version #1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/690,494
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MacLeod, Janet M.
REGISTRATION NUMBER: 35,263
REFERENCE/DOCKET NUMBER: 29545-A-PCT/USA-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2597
TELEFAX: 212-765-2519

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 120 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: misc.feature

LOCATION: 1..120

US-08-690-494-26

Query Match

Best Local Similarity 66.7%; Score 22; DB 2; Length 120;
Matches 33; Conservative 97.1%; Pred. No. 0.27;

Mismatches 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGACGACCATATGCG 33

Db 64 CGGTGTACTCACCAGTCCGACGACCATATGCG 97

RESULT 3

PCT-US95-07671-26

Sequence 26 Application PC/TUS9507671

GENERAL INFORMATION:

APPLICANT: Zhang, David Y.
TITLE OF INVENTION: LIGATION-DEPENDENT AMPLIFICATION FOR THE
NUMBER OF SEQUENCES: 31
DETECTION OF INFECTIOUS PATHOGENS AND ABNORMAL GENES
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
ZIP: 10112-0228

COUNTRY: USA

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/07671

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Seide, Rochelle K

REGISTRATION NUMBER: 32,300

REFERENCE/DOCKET NUMBER: 29545-A-PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-408-2626

TELEFAX: 212-765-2519

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 120 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: misc.feature

LOCATION: 1..120

PCT-US95-07671-26

Query Match

Best Local Similarity 66.7%; Score 22; DB 5; Length 120;
Matches 33; Conservative 97.1%; Pred. No. 0.27;

Mismatches 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGACGACCATATGCG 33

Db 64 CGGTGTACTCACCAGTCCGACGACCATATGCG 97

RESULT 4

US-08-690-495-18

Sequence 18 Application US/08690495

GENERAL INFORMATION:

APPLICANT: Zhang, David Y., Brandwein, Margaret
TITLE OF INVENTION: NUCLEIC ACID AMPLIFICATION METHOD:
NUMBER OF SEQUENCES: 42
HYBRIDIZATION SIGNAL AMPLIFICATION METHOD (HSAM)
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
ZIP: 10112-0228

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Fasteq Version #1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/690,495

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MacLeod, Janet M.
REGISTRATION NUMBER: 35,263
REFERENCE/DOCKET NUMBER: 29545-A-PCT/USA-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2597
TELEFAX: 212-765-2519

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 145 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..145
US-08-690-495-18

Query Match 66.7%; Score 22; DB 2; Length 145;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 33
DB 72 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 105

RESULT 5
US-08-690-494-18
Sequence 18, Application US/08690494
Patent No. 5942391

GENERAL INFORMATION:

APPLICANT: Zhang, David Y., Brandwein, Margaret
TITLE OF INVENTION: NUCLEIC ACID AMPLIFICATION METHOD:
TITLE OF INVENTION: HYBRIDIZATION SIGNAL AMPLIFICATION METHOD (HSAM)
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:

ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY

COUNTRY: USA

ZIP: 10112-0228

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: FASEQ Version #1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/690,494

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MacLeod, Janet M.

REGISTRATION NUMBER: 35,263

REFERENCE/DOCKET NUMBER: 29545-A-PCT/USA-B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-408-2597

TELEFAX: 212-765-2519

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 145 base pairs

TYPE: nucleic acid

STRANDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..145

US-08-690-494-18

Query Match 66.7%; Score 22; DB 2; Length 145;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 33
DB 72 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 105

RESULT 6
PCT-US95-07671-18
Sequence 18, Application PC/TUS9507671
GENERAL INFORMATION:
APPLICANT: Zhang, David Y.

TITLE OF INVENTION: LIGATION-DEPENDENT AMPLIFICATION FOR THE
DETECTION OF INFECTIOUS PATHOGENS AND ABNORMAL GENES

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10112-0228

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/07671

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Seide, Rochelle K.

REGISTRATION NUMBER: 32,300

REFERENCE/DOCKET NUMBER: 29545-A-PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-408-2626

TELEFAX: 212-765-2519

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 145 base pairs

TYPE: nucleic acid

STRANDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..145

PCT-US95-07671-18

Query Match 66.7%; Score 22; DB 5; Length 145;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 33
DB 72 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 105

RESULT 7
US-08-256-568B-55/C
Sequence 55, Application US/08256568B
Patent No. 5846704

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;

APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO

TITLE OF INVENTION: PROCESS FOR TYPING OF HCV

TITLE OF INVENTION: ISOLATES

NUMBER OF SEQUENCES: 97

CORRESPONDENCE ADDRESS:

ADDRESSEE: BIERMAN & MUSERLIAN

STREET: 600 THIRD AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10016

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCIT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/256,568B

FILING DATE: 18-JUL-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410.004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
CLONE: jp62
POSITION IN GENOME:
MAP POSITION: 5' untranslated region
US-08-256-568B-55

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

DB 1 CGGTGACTCACC-GTTCGCGAGACCATATGCG 33
73 CGGTGACTCACC-GTTCGCGAGACCATATGCG 40

RESULT 8
US-08-256-568B-56/c
Sequence 56, Application US/08256568B
Patent No. 5846704
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HCV
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992

ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410.004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
CLONE: gp81
POSITION IN GENOME:
MAP POSITION: 5' untranslated region
US-08-256-568B-56

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

DB 1 CGGTGACTCACC-GTTCGCGAGACCATATGCG 33
73 CGGTGACTCACC-GTTCGCGAGACCATATGCG 40

RESULT 9
US-08-256-568B-57/c
Sequence 57, Application US/08256568B
Patent No. 5846704
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HCV
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410.004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:

LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: b156
POSITION IN GENOME:
MAP POSITION: 5' untranslated region
US-08-256-568B-57

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 33
Db 73 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 40

RESULT 10
US-08-256-568B-58/c
Sequence 58, Application US/08256568B
Patent No. 5846704
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HCV
NUMBER OF INVENTION: ISOLATES
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410,004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: bu79
POSITION IN GENOME:
MAP POSITION: 5' untranslated region

US-08-256-568B-58

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 33
Db 73 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 40

RESULT 11
US-08-256-568B-61/c
Sequence 61, Application US/08256568B
Patent No. 5846704
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HCV
NUMBER OF INVENTION: ISOLATES
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410,004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: be82 (also referred to as be99)
POSITION IN GENOME:
MAP POSITION: 5' untranslated region
US-08-256-568B-61

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 33
Db 73 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 40

RESULT 12

US-08-256-568B-62/c
Sequence 62, Application US/08256568B
Patent No. 5846704
GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HCV
TITLE OF INVENTION: ISOLATES
NUMBER OF SEQUENCES: 97

CORRESPONDENCE ADDRESS:

ADDRESS: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992

ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410.004
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 62:

SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: be90

POSITION IN GENOME:
MAP POSITION: 5' untranslated region

US-08-256-568B-62

Query Match:

Best Local Similarity 66.7%; Score 22; DB 2; Length 177;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 73 CGGTGACTACACCGTCCGACGACCACTATGCC 33
CGGTGACTACACCGTCCGACGACCACTATGCC 40

RESULT 13

US-08-256-568B-63/c
Sequence 63, Application US/08256568B
Patent No. 5846704
GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO

TITLE OF INVENTION: PROCESS FOR TYPING OF HCV
TITLE OF INVENTION: ISOLATES
NUMBER OF SEQUENCES: 97

CORRESPONDENCE ADDRESS:
ADDRESS: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992

ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410.004
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 63:

SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: be91

POSITION IN GENOME:
MAP POSITION: 5' untranslated region

US-08-256-568B-63

Query Match:
Best Local Similarity 66.7%; Score 22; DB 2; Length 177;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 73 CGGTGACTACACCGTCCGACGACCACTATGCC 33
CGGTGACTACACCGTCCGACGACCACTATGCC 40

RESULT 14

US-08-256-568B-64/c
Sequence 64, Application US/08256568B
Patent No. 5846704
GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HCV
TITLE OF INVENTION: ISOLATES
NUMBER OF SEQUENCES: 97

CORRESPONDENCE ADDRESS:
ADDRESS: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992

ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410.004
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 64:

SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: be91

POSITION IN GENOME:
MAP POSITION: 5' untranslated region

US-08-256-568B-64

ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410.004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: be92
POSITION IN GENOME:
MAP POSITION: 5' untranslated region
US-08-256-568B-64

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGACTCACC-GTTCGCGAGACCACTATGCG 33
DB 73 CGGTGACTCACC-GTTCGCGAGACCACTATGCG 40

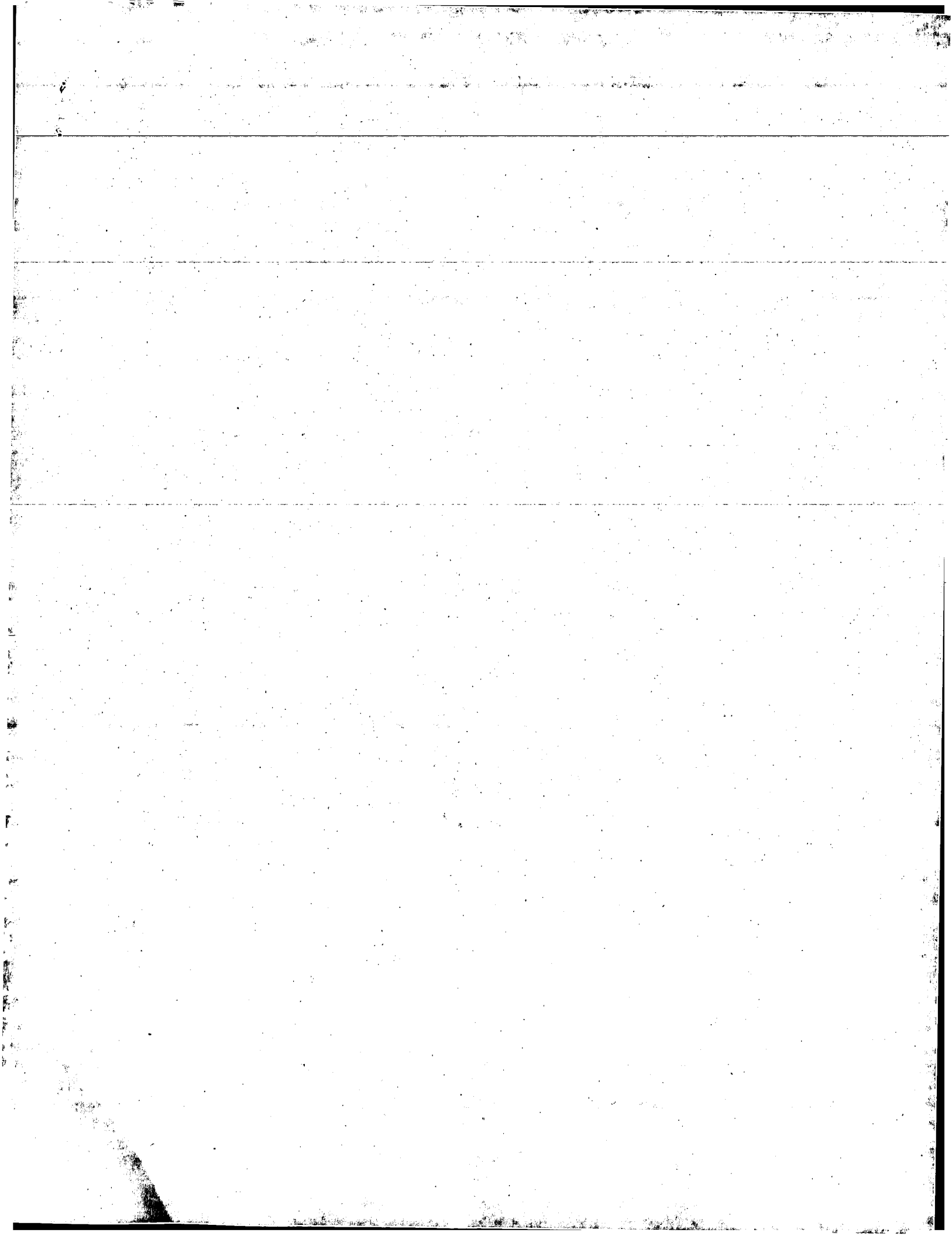
RESULT 15
US-08-256-568B-65/C
Sequence 65, Application US/08256568B
Patent No. 5846704
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HCY
TITLE OF INVENTION: ISOLATES
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410.004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: be93
POSITION IN GENOME:
MAP POSITION: 5' untranslated region
US-08-256-568B-65

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGACTCACC-GTTCGCGAGACCACTATGCG 33
DB 73 CGGTGACTCACC-GTTCGCGAGACCACTATGCG 40

Search completed: July 10, 2003, 20:27:12
Job time: 15.7573 secs



/db.xref="taxon:8364"
 /clone="Tneu13m07"
 /clone.lib="XGC-neurula"
 /dev_stage="neurula"
 /lab_host="Escherichia coli DH10B"
 /note="Vector: pCS107; Site.1: EcoRI; Site.2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."
 104 a 171 c 175 g 84 t

BASE COUNT
 ORIGIN
 Query Match 14.5%; Score 35; DB 9; Length 534;
 Best Local Similarity 50.3%; Pred. No. 19;
 Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 17 CATGCGCTAGTATGATGTCGTGACGCTCCAGAGACCCCTCCCGGAGAGCCATAG 76
 Db 313 CATGGCGGTACTACAGGCACAGGCGATCTCAGAGGGGACGAGAGACTGGCGCGGGA 254
 QY 77 TGTCTCGCGAAGCGGTGATGATACCGGAATTGCGAGAGAGCGGCTCTTCTTGAT 136
 Db 253 TGTCTCGGAGAGAGATGCTGAGACGCTGATAGAGCGGGCTGCGGGGTATTTCACGCGCT 194
 QY 137 CAACCGCTCAATGCTGAGATTTGGGCGTCCCGCGAGACTGCTAGC 187
 Db 193 CTAGCCCTTACACTCCCGGGGATTCAGCCTACTGACGCGGCTTGTGGC 143

RESULT 2
 AL647223/c 557 bp mRNA linear EST 13-DEC-2001
 LOCUS AL647223 XGC-gastrula Silurana tropicalis cDNA clone Tgas026a22 5'
 DEFINITION mRNA sequence.
 VERSION AL647223.1 GI:17655336
 KEYWORDS EST.
 SOURCE Western clawed frog
 ORGANISM Silurana tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Silurana.
 1 (bases 1 to 557)
 Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
 Sanger Xenopus tropicalis EST project 2001 (10_2001)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Huckle E
 Sanger Centre
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trop@sanger.ac.uk
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS_SEQUENCE.ID: Tgas026a22.sp6
 Sequencing primer: Sp6
 This sequence is from a Xenopus Gene Collection (XGC) library
 constructed by Aaron M. Zorn.
 Location/Qualifiers
 1..557

FEATURES
 source
 /organism="Silurana tropicalis"
 /db.xref="taxon:8364"
 /clone="Tgas026a22"
 /clone.lib="XGC-gastrula"
 /dev_stage="gastrula (stages 10.5-13 mixed)"
 /lab_host="Escherichia coli XL1-blue"
 /note="Vector: pCS107; Site.1: EcoRI; Site.2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."
 105 a 179 c 186 g 87 t

BASE COUNT
 ORIGIN
 Query Match 14.5%; Score 35; DB 9; Length 557;
 Best Local Similarity 50.3%; Pred. No. 20;

Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
 QY 17 CATGCGCTAGTATGATGTCGTGACGCTCCAGAGACCCCTCCCGGAGAGCCATAG 76
 Db 313 CATGGCGGTACTACAGGCACAGGCGATCTCAGAGGGGACGAGAGACTGGCGCGGGA 254
 QY 77 TGTCTCGCGAAGCGGTGATGATACCGGAATTGCGAGAGAGCGGCTCTTCTTGAT 136
 Db 253 TGTCTCGGAGAGAGATGCTGAGACGCTGATAGAGCGGGCTGCGGGGTATTTCACGCGCT 194
 QY 137 CAACCGCTCAATGCTGAGATTTGGGCGTCCCGCGAGACTGCTAGC 187
 Db 193 CTAGCCCTTACACTCCCGGGGATTCAGCCTACTGACGCGGCTTGTGGC 143

RESULT 3
 AL636391/c 572 bp mRNA linear EST 12-DEC-2001
 LOCUS AL636391 XGC-neurula Silurana tropicalis cDNA clone Tneu021f04 5'
 DEFINITION mRNA sequence.
 VERSION AL636391.1 GI:16788370
 KEYWORDS EST.
 SOURCE Western clawed frog
 ORGANISM Silurana tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Silurana.
 1 (bases 1 to 572)
 Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
 Sanger Xenopus tropicalis EST project 2001 (10_2001)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Huckle E
 Sanger Centre
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trop@sanger.ac.uk
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS_SEQUENCE.ID: Tneu021f04.sp6
 Sequencing primer: Sp6
 This sequence is from a Xenopus Gene Collection (XGC) library
 constructed by Aaron M. Zorn.
 Location/Qualifiers
 1..572

FEATURES
 source
 /organism="Silurana tropicalis"
 /db.xref="taxon:8364"
 /clone="Tneu021f04"
 /clone.lib="XGC-neurula"
 /dev_stage="neurula"
 /lab_host="Escherichia coli DH10B"
 /note="Vector: pCS107; Site.1: EcoRI; Site.2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."
 98 a 180 c 199 g 95 t

BASE COUNT
 ORIGIN
 Query Match 14.5%; Score 35; DB 9; Length 572;
 Best Local Similarity 50.3%; Pred. No. 20;
 Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 17 CATGCGCTAGTATGATGTCGTGACGCTCCAGAGACCCCTCCCGGAGAGCCATAG 76
 Db 268 CATGGCGGTACTACAGGCACAGGCGATCTCAGAGGGGACGAGAGACTGGCGCGGGA 209
 QY 77 TGTCTCGCGAAGCGGTGATGATACCGGAATTGCGAGAGAGCGGCTCTTCTTGAT 136
 Db 208 TGTCTCGGAGAGAGATGCTGAGACGCTGATAGAGCGGGCTGCGGGGTATTTCACGCGCT 149
 QY 137 CAACCGCTCAATGCTGAGATTTGGGCGTCCCGCGAGACTGCTAGC 187
 Db 148 CTAGCCCTTACACTCCCGGGGATTCAGCCTACTGACGCGGCTTGTGGC 98

RESULT 4
AL784104/c 577 bp mRNA linear EST 25-JUN-2002
LOCUS AL784104 XGC-gastrula Silurana tropicalis cDNA clone Tgas073h08 5',
DEFINITION mRNA sequence.
ACCESSION AL784104
KEYWORDS GI:21569808
VERSION EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesodactyla; Pipridae; Pipridae;
Xenopodinae; Silurana.
REFERENCE 1 (bases 1 to 577)
AUTHORS Taylor, R., Ashurst, J.L., Croning, M.D.R., Zorn, A.M. and Rogers, J.
TITLE Sanger Xenopus tropicalis EST project 2002
JOURNAL Unpublished (2001)
COMMENT Contact: Taylor R
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: tropesanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: Tgas073h08.plksp6
Sequencing primer: PKSP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron W. Zorn.
location/Qualifiers
FEATURES
source
1. 577
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="Tgas073h08"
/dev_stage="gastrula (stages 10.5-13 mixed)"
/lab_host="Escherichia coli XL1-blue"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."
BASE COUNT 108 a 184 c 194 g 91 t
ORIGIN
Query Match 14.5%; Score 35; DB 9; Length 577;
Best Local Similarity 50.3%; Pred. No. 20;
Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
OY 17 CATGCGCTAGTATGAGTGTGCTGCGATCCAGACCCCTCCGGGAGAGCATTAG 76
DB 321 CATGGCGTACTCAGCGACAGGCGATCCAGACGAGGAGAGAGCTGGGCGACAGCGGA 262
OY 77 TGTCTCTGCGAAGCGGTGAGTACACCGGAATTGCCAGAGACGCGGTCCTTTCTTGAT 136
DB 261 TGCTCCGAGAAAGATGCTGAACGCTGATAGACGGGCTGCGGGGATTTGACAGGCT 202
OY 137 CAACCCGCTCAATGCTCGAGATTGGGCGTCCCGCGGAGACTGCTAGC 187
DB 201 CTAGCCCTTACATCCCGGGGATTGACCTTACGACCGCGGCTTGTGGC 151

RESULT 5
BE249938/c 928 bp mRNA linear EST 13-JUL-2000
LOCUS BE249938 600942989F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:2959391 5',
DEFINITION mRNA sequence.
ACCESSION BE249938
KEYWORDS GI:9120041
VERSION EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 928)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at: image.llnl.gov
Plate: LCM51 row: a column: 24
High quality sequence start: 2
High quality sequence stop: 758.
location/Qualifiers
FEATURES
source
1. 928
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2959391"
/clone="IB-NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 223 a 230 c 319 g 156 t
ORIGIN
Query Match 14.3%; Score 34.4; DB 10; Length 928;
Best Local Similarity 48.5%; Pred. No. 33;
Matches 95; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
OY 40 GCAGCTTCAGAGACCCCTCCCGGAGAGCCATTAGTGTCTGCGAAGCGGTGAGTAC 99
DB 819 GACACTCCAGCGCTCCATCCATGAGCCGCCAGGAGCCCTTGAGCCCTCA 760
OY 100 ACCGGAATTGCGAGACGACCGGCTTCTTGATCAACCCGCTCAATGCTGGAGAT 159
DB 759 TCCAGCCGCGACCGGATGTCGGTTCATCTAGAGATCACTTTGGCCGCGAGTAT 700
OY 160 TTGGGCGTCCCGCCGCGAGACTGCTAGTGTGAGTGTGCGGAAAGGCTTGTGG 219
DB 699 CCCCCTGTTCCAGACCGAGGCTCGAGCTCTCATACTTATGCTACATGCTCTCAGC 640
OY 220 TACTGCTGATAGGCT 235
DB 639 CTCAGCCCGCTGCGGT 624

RESULT 6
AL563471 952 bp mRNA linear EST 16-FEB-2001
LOCUS AL563471 LTI.NF1001.NBC4 Homo sapiens cDNA clone CSDD006YF10 3
DEFINITION prime, mRNA sequence.
ACCESSION AL563471
KEYWORDS GI:12912895
VERSION EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 952)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqlref@genoscope.cns.fr, web : www.genoscope.cns.fr.

100

Matches	58	Conservative	0	Mismatches	42	Indels	0	Gaps	0
QY	63	CGGAGAGACCACTAGTGTCTCCGGAACCGGTGATACACCGGAATTGCGACAGACCGG	122						
Db	243	CAGGGGCACTCCAGCACTGCTGTGACTCTGTATGGGGGCACTGGGGTGTCTGGAGCTCAGG	184						
QY	123	GTCCTTTCCTGATCAACCCGCTCAATGCTCGAGAGATTG	162						
Db	183	GCCCATCACTGTGCTGCACCTCCCTCCACGCGCTGGGGGAATG	144						
RESULT 13									
CNS0091P									
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC #								
LOCUS	BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit								
FLY)	fly), genomic survey sequence.								
ACCESSION	AL053013								
VERSION	AL053013.1								
KEYWORDS	GSS.								
ORGANISM	Drosophila melanogaster.								
SOURCE	Drosophila melanogaster.								
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;								
AUTHORS	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;								
TITLE	Ephydroidea; Drosophilidae; Drosophila.								
JOURNAL	1 (bases 1 to 925)								
COMMENT	Genoscope.								
	Direct Submission								
	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;								
	Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr								
	- Web : www.genoscope.cns.fr)								
	Determination of this BAC-end sequence was carried out as part of a								
	collaboration with the Berkeley Drosophila Genome Project (BDGP).								
	The BDGP is constructing a physical map of the Drosophila								
	melanogaster genome using these BACs. For further information								
	please see http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library								
	melanogaster BAC library was prepared by Kazutoyo Osoegawa and								
	Aaron Mammosser in Pieter de Jong's laboratory in the Department of								
	Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,								
	NY. The library is named RPCI-98 and was constructed by partial								
	ECORI digestion of Drosophila DNA provided by the BDGP from the								
	isogenic strain Y2: cn bw ap, the same strain used for the BDGP's								
	PI and EST libraries. A more detailed description of the library								
	and how to order individual BAC clones, the entire library, or								
	filters for hybridization from the BACPAC Resource Center can be								
	found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .								
FEATURES									
source	location/Qualifiers								
	1..925								
	/organism="Drosophila melanogaster"								
	/db_xref="taxon:7227"								
	/clone="BACR19D16"								
	/clone_lib="RPCI-98"								
	/note="end : TET3"								
BASE COUNT	120 a 61 c 61 g 172 t 511 others								
ORIGIN									
Query Match	13.6% ; Score 32.8 ; DB 17 ; Length 925 ;								
Best Local Similarity	13.7% ; Pred. No. 91 ;								
Matches	29 ; Conservative 98 ; Mismatches 85 ; Indels 0 ; Gaps 0 ;								
QY	1 GCAGAGACGCTGACCATGCGCTAGTAGTGTCTGTCGACCGCTCCAGACCCCCCT 60								
Db	707 SSSVSAGSKSTBSGSSSSSSSSSSSTSSBBSCTSSSSSSSSSSSSSSCTCCCS 766								
QY	61 CCGCGGAGACCATGCTGTCTGCGGACACCGCTAGTACACCGGAATTGCGACGAC 120								
Db	767 YSYSSSTSSSTSTGSTSSSSSSSVGSSSSSDSTSTCCCTCCCTCCTYBMCCYST 826								
QY	121 GGGTCCTTCTTGATCATACCCGCTCATGCTGTGAGATTTGGCGGTGCGGAGAC 180								
Db	827 SCGGSSSSSGGCGYTKCGCGGCGSSSTTGMGTSSACSSSSSSSSSVSSSKSSASS 886								
QY	181 TGTAGCCGAGTAGTGTGGGTGCGGAAAGC 212								

DB	SVSSSGSSGVSNSSSASKSSSSGVS	SVSSGSS	918
RESULT 14	BF306437	1331 bp	MRNA
LOCUS	60189332221 NIH_MGC_17	Homo sapiens	CDNA clone IMAGE:4138900 5',
DEFINITION	MRNA sequence.		
ACCESSION	BF306437		
VERSION	BF306437.1	GI:11253532	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1. (bases 1 to 1331)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@bbs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN). DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: image.lnl.gov Plate: LCM048 row: 1 column: 05 High quality sequence stop: 642. Location/Qualifiers 1. 1331 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4138900" /clone_lib="NIH_MGC_17" /tissue_type="rhabdomyosarcoma" /lab_host="DH10B (phage-resistant)" /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGCAG(C). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."		
BASE COUNT	419 a 380 c 382 g 147 t	3 others	
ORIGIN			
Query Match	13.6%; Score 33.8; DB 12; Length 1331;		
Best Local Similarity	58.0%; Pred. No. 1e+02;		
Matches	58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;		
QY	63 CGGAGACCACTAGTGTGTCGGGAACCGGTGATGATACACCGGAATTGCCAGAGCAGCCG	122	
DB	224 CAGGGCAATCCAGCACTGCTGTGACTCTGTATGGGGCCACTGGGGCTTGTGGAGCCTCAG	165	
QY	123 GTCCTTTCTTGATCAACCGCTCAATGCTGGAGATTG	162	
DB	164 GGCATCACTGCTGCTCCGCTCCACGGCTGGGGGAATG	125	
RESULT 15	BC014080/c	2670 bp	MRNA
LOCUS	BC014080		
DEFINITION	Homo sapiens, clone IMAGE:4309173.		
ACCESSION	BC014080		
VERSION	BC014080.1	GI:17315140	
KEYWORDS	HTC.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		

Fri Jul 11 15:29:33 2003

us-10-087-631b-10.rst

REFERENCE 1 (bases 1 to 2670)
 AUTHORS Strausberg R.
 TITLE Direct Submission
 JOURNAL Submitted (10-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Louis Staudt
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web Site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nrl.nih.gov
 Shevchenko, Y., Wetherby, R.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C., McQuell, J., Pearson, R., Snyder, B., Stantrop, S., Thomas, P.J., Tlonsong, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAL Plate: 26 Row: 1 Column: 4
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis
 This clone has the following problem: Incomplete processing

FEATURES
 source
 Location/Qualifiers
 1..2670
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4309173"
 /tissue_type="Primary B-Cells from Tonsils"
 /clone_lib="NIH-MGC_48"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"

BASE COUNT 657 a 742 c 774 g 497 t
 ORIGIN

Query Match 13.6%; Score 32.8; DB 11; Length 2670;
 Best Local Similarity 58.0%; Pred. No. 1.2e+02;
 Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY	63	CGGGAGAGCATAGTGTGCGGACCGGTGATACACCGAAATTGCCAGACGACCGG	122
DB	776	CAGGGCATCCACACACTGTCTGTGATCTGTAGGGCCACCTGGGTTGCTGGACGTACG	717
QY	123	GTCTCTTCTTGATCAACCGCTCATGCTGGAGATTGG	162
DB	716	GGCATTCACTGCTGCTCCTCCACGCGTGGGGAATG	677

Search completed: July 11, 2003, 02:26:19
 Job time : 4340.13 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 2088.31 Seconds

(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631B-10

Perfect score: 241
Sequence: 1 gcagaagagcgtctagccatg.....ctgcctgataggtgcttgc 241

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenBank1:
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	241	100.0	244	6 A98437	A98437 Sequence 21
2	241	100.0	254	14 D31724	D31724 Hepatitis C
3	241	100.0	256	14 AF217300	AF217300 Hepatitis C
4	241	100.0	279	14 HPCBCU5	L34388 Hepatitis C
5	241	100.0	305	6 AX398187	AX398187 Sequence
6	241	100.0	321	14 HPCRNAD3	M84842 Hepatitis C
7	241	100.0	321	14 HPCRNAD5	M84841 Hepatitis C
8	241	100.0	321	14 HPCRNAD4	M84859 Hepatitis C
9	241	100.0	321	14 HPCRNAD6	M84859 Hepatitis C
10	241	100.0	321	14 HPCRNAD7	M84859 Hepatitis C
11	241	100.0	321	14 HPCRNAD8	M84859 Hepatitis C
12	241	100.0	321	14 HPCRNAD9	M84859 Hepatitis C
13	241	100.0	321	14 HPCRNAD10	M84859 Hepatitis C
14	241	100.0	321	14 HPCRNAD11	M84859 Hepatitis C
15	241	100.0	321	14 HPCRNAD12	M84859 Hepatitis C
16	241	100.0	321	14 HPCRNAD13	M84859 Hepatitis C
17	241	100.0	321	14 HPCRNAD14	M84859 Hepatitis C
18	241	100.0	321	14 HPCRNAD15	M84859 Hepatitis C
19	241	100.0	321	14 HPCRNAD16	M84859 Hepatitis C
20	241	100.0	321	6 AR027784	AR027784 Sequence
21	241	100.0	328	6 AX419905	AX419905 Sequence
22	241	100.0	341	6 AR119855	AR119855 Sequence
23	241	100.0	341	6 AR119856	AR119856 Sequence
24	241	100.0	341	6 AR153763	AR153763 Sequence
25	241	100.0	341	14 HPC5NR2	D31602 Human hepat
26	241	100.0	342	6 HPC5NTR	L40552 Hepatitis C
27	241	100.0	366	6 AR095001	AR095001 Sequence
28	241	100.0	395	6 AX282476	AX282476 Sequence
29	241	100.0	483	6 HCU05028	U05028 Hepatitis C
30	241	100.0	483	6 E06161	E06161 cDNA encodi
31	241	100.0	587	6 E06357	E06357 cDNA encodi
32	241	100.0	686	6 E50749	E50749 Vector expr
33	241	100.0	686	6 AR142355	AR142355 Sequence
34	241	100.0	686	6 AR167008	AR167008 Sequence
35	241	100.0	703	6 AR210663	AR210663 Sequence
36	241	100.0	780	6 E50750	E50750 Vector expr
37	241	100.0	866	6 AR095006	AR095006 Sequence
38	241	100.0	923	14 HCU94722	U94722 Hepatitis C
39	241	100.0	1499	6 AR153761	AR153761 Sequence
40	241	100.0	1499	6 AR005071	AR005071 Sequence
41	241	100.0	1499	6 AR064510	AR064510 Sequence
42	241	100.0	1499	6 AR091903	AR091903 Sequence
43	241	100.0	1554	6 AR145552	AR145552 Sequence
44	241	100.0	1554	6 E06166	E06166 cDNA encodi
45	241	100.0	1580	6 E06362	E06362 cDNA encodi
			14	6 DB8470	DB8470 Hepatitis C

ALIGNMENTS

RESULT 1
A98437
LOCUS A98437
DEFINITION Sequence 21 from Patent WO9912948.
ACCESSION A98437
VERSION A98437.1 GI:6781538
KEYWORDS
SOURCE
' ORGANISM
REFERENCE 1 (bases 1 to 244)
AUTHORS
TITLE
JOURNAL

Protein-coated polyribonucleic acids, method for the production thereof, and use of the same
Patent: WO 9912948-A 21 18-MAR-1999;

FEATURES
SOURCE
LANDT OLFERT (DE)
Location/Qualifiers
1..244
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT
45 a 67 c 81 g 51 t

Query Match
Best Local Similarity 100.0%; Score 241; DB 6; Length 244;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCATGCGGCTTGTAGTGTGTCGACGCTCCAGAGACCCCTT 60
DB 1 GCAGAAAGCGTCTAGCATGCGGCTTGTAGTGTGTCGACGCTCCAGAGACCCCTT 60
QY 61 CCCGGAGAGCCATAGTGTGTCGGAACCGGTGAGTACCGGAATTCAGAGAC 120
DB 61 CCCGGAGAGCCATAGTGTGTCGGAACCGGTGAGTACCGGAATTCAGAGAC 120
QY 121 GGGTCTTCTTGGATCAACCCGCTCAATGCTGAGATTGGGGGTGCGCGAGAC 180
DB 121 GGGTCTTCTTGGATCAACCCGCTCAATGCTGAGATTGGGGGTGCGCGAGAC 180
QY 181 TCGTAGCCGAGTAGTGTGTCGCGAAGGCGCTTGTGCTGCTGATAGGTCCTTG 240
DB 181 TCGTAGCCGAGTAGTGTGTCGCGAAGGCGCTTGTGCTGCTGATAGGTCCTTG 240
QY 241 C 241
DB 241 C 241

RESULT 2
D31724 254 bp RNA linear VRL 09-SEP-1997
LOCUS Hepatitis C virus genomic RNA, 5'UTR, clone:BUG.
DEFINITION D31724
ACCESSION D31724.1 GI:2388524
VERSION D31724.1 GI:2388524
KEYWORDS
SOURCE Hepatitis C virus cDNA to genomic RNA, clone:BUG.
ORGANISM Hepatitis C virus
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE
1 (sites)
AUTHORS Ohta,K., Mizokami,M., Ohno,T., Suzuki,K., Orito,E., Ina,Y.,
Lau,J.Y. and Gojobori,T.
TITLE Classification of hepatitis C virus into major types and subtypes
based on molecular evolutionary analysis
JOURNAL Virus Res. 36 (2-3), 201-214 (1995)
MEDLINE 95381614
REFERENCE 2 (bases 1 to 254)
AUTHORS Suzuki,K.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-1994) Kaoru Suzuki, Nagoya City University
Medical School, Second Department of Internal Medicine, 1-Kawasumi,
Mizuh, Mizuh-ku, Nagoya, Aichi 466, Japan (Tel:052-853-8748,
Fax:052-852-0849)

FEATURES
SOURCE
Location/Qualifiers
1..254
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
/clone="BUG"

5'UTR
BASE COUNT
46 a 72 c 82 g 54 t

Query Match
Best Local Similarity 100.0%; Score 241; DB 14; Length 254;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCATGCGGCTTGTAGTGTGTCGACGCTCCAGAGACCCCTT 60
DB 1 GCAGAAAGCGTCTAGCATGCGGCTTGTAGTGTGTCGACGCTCCAGAGACCCCTT 60

DB 6 GCAGAAAGCGTCTAGCATGCGGCTTGTAGTGTGTCGACGCTCCAGAGACCCCTT 65
QY 61 CCCGGAGAGCCATAGTGTGTCGGAACCGGTGAGTACCGGAATTCAGAGAC 120
DB 66 CCCGGAGAGCCATAGTGTGTCGGAACCGGTGAGTACCGGAATTCAGAGAC 125
QY 121 GGGTCTTCTTGGATCAACCCGCTCAATGCTGAGATTGGGGGTGCGCGAGAC 180
DB 126 GGGTCTTCTTGGATCAACCCGCTCAATGCTGAGATTGGGGGTGCGCGAGAC 185
QY 181 TCGTAGCCGAGTAGTGTGTCGCGAAGGCGCTTGTGCTGCTGATAGGTCCTTG 240
DB 186 TCGTAGCCGAGTAGTGTGTCGCGAAGGCGCTTGTGCTGCTGATAGGTCCTTG 245
QY 241 C 241
DB 246 C 246

RESULT 3
AF217300 256 bp RNA linear VRL 23-MAR-2000
LOCUS AF217300
DEFINITION Hepatitis C virus clone Sot12 5'UTR sequence.
ACCESSION AF217300
VERSION AF217300.1 GI:7288521
KEYWORDS
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE
1 (bases 1 to 256)
AUTHORS Green,S.M. and Hadfield,S.J.
TITLE Direct Submission
JOURNAL Submitted (20-DEC-1999) Microbiology, Public Health Laboratory
Service, Southampton General Hospital, Southampton, Hampshire SO16
6YD, United Kingdom

FEATURES
SOURCE
Location/Qualifiers
1..256
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
/clone="Sot12"
/note="genotype: 1b"

5'UTR
BASE COUNT
46 a 73 c 85 g 52 t

Query Match
Best Local Similarity 100.0%; Score 241; DB 14; Length 256;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCATGCGGCTTGTAGTGTGTCGACGCTCCAGAGACCCCTT 60
DB 4 GCAGAAAGCGTCTAGCATGCGGCTTGTAGTGTGTCGACGCTCCAGAGACCCCTT 63
QY 61 CCCGGAGAGCCATAGTGTGTCGGAACCGGTGAGTACCGGAATTCAGAGAC 120
DB 64 CCCGGAGAGCCATAGTGTGTCGGAACCGGTGAGTACCGGAATTCAGAGAC 123
QY 121 GGGTCTTCTTGGATCAACCCGCTCAATGCTGAGATTGGGGGTGCGCGAGAC 180
DB 124 GGGTCTTCTTGGATCAACCCGCTCAATGCTGAGATTGGGGGTGCGCGAGAC 183
QY 181 TCGTAGCCGAGTAGTGTGTCGCGAAGGCGCTTGTGCTGCTGATAGGTCCTTG 240
DB 184 TCGTAGCCGAGTAGTGTGTCGCGAAGGCGCTTGTGCTGCTGATAGGTCCTTG 243
QY 241 C 241
DB 244 C 244

RESULT 4
HPCBCUS

LOCUS HPCBCU5 279 bp DNA linear VRL 16-MAY-1995
 DEFINITION Hepatitis C virus 5' non-coding region.
 ACCESSION L34388
 VERSION L34388.1 GI:511628
 KEYWORDS noncoding region.
 SOURCE Hepatitis C virus DNA.
 ORGANISM Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE 1 (bases 1 to 279)
 Altamirano, M., Delaney, A., Wong, A., Marosteenmak, J. and P. D.
 Identification of hepatitis C virus genotypes among hospitalized
 patients in British Columbia, Canada
 J. Infect. Dis. 171 (4), 1034-1038 (1995)
 MEDLINE 9521959
 PUBMED 7706784

FEATURES
 source
 1. 279
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"

BASE COUNT 49 a 79 c 92 g 59 t
 ORIGIN

Query Match 100.0%; Score 241; DB 14; Length 279;
 Best Local Similarity 100.0%; Pred. No. 6.5e-60;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAACGCTCTACCCATGCGGCTTATGATGATGCTGTCGACGCTCCAGAGACCCCT 60
 DB 9 GCAGAAACGCTCTACCCATGCGGCTTATGATGATGCTGTCGACGCTCCAGAGACCCCT 68
 OY 61 CCCGGAGAGCCATGCTGCTGCGGAGACCGGTGATGACCGGATGCCAGAGACGACC 120
 DB 69 CCCGGAGAGCCATGCTGCTGCGGAGACCGGTGATGACCGGATGCCAGAGACGACC 128
 OY 121 GGGTCCTTCTTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 180
 DB 129 GGGTCCTTCTTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 188
 OY 181 TGTACGCGAGTATGTTGGTGGCGGAAAGGCTTGTGTACTGCTGATAGAGGTTG 240
 DB 189 TGTACGCGAGTATGTTGGTGGCGGAAAGGCTTGTGTACTGCTGATAGAGGTTG 248
 OY 241 C 241
 DB 249 C 249

RESULT 5
 AX398187 305 bp DNA linear PAT 27-MAY-2002
 LOCUS AX398187
 DEFINITION Sequence 64 from Patent WO0220837.
 ACCESSION AX398187
 VERSION AX398187.1 GI:21261002
 KEYWORDS
 SOURCE Hepatitis C virus.
 ORGANISM Hepatitis C virus.
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE 1
 Ronaghi, M., Ekstrom, B. and Pourmand, N.
 Method
 Patent: WO 0220837-A 64 14-MAR-2002;
 Pyrosequencing AB (SE); The Board of Trustees of The Leland
 Stanford Junior University (US)
 Location/Qualifiers
 1. 305
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"

BASE COUNT 56 a 87 c 98 g 64 t
 ORIGIN

Query Match 100.0%; Score 241; DB 6; Length 305;

Best Local Similarity 100.0%; Pred. No. 6.5e-60;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAACGCTCTACCCATGCGGCTTATGATGATGCTGTCGACGCTCCAGAGACCCCT 60
 DB 26 GCAGAAACGCTCTACCCATGCGGCTTATGATGATGCTGTCGACGCTCCAGAGACCCCT 85
 OY 61 CCCGGAGAGCCATGCTGCTGCGGAGACCGGTGATGACCGGATGCCAGAGACGACC 120
 DB 86 CCCGGAGAGCCATGCTGCTGCGGAGACCGGTGATGACCGGATGCCAGAGACGACC 145
 OY 121 GGGTCCTTCTTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 180
 DB 146 GGGTCCTTCTTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 205
 OY 181 TGTACGCGAGTATGTTGGTGGCGGAAAGGCTTGTGTACTGCTGATAGAGGTTG 240
 DB 206 TGTACGCGAGTATGTTGGTGGCGGAAAGGCTTGTGTACTGCTGATAGAGGTTG 265
 OY 241 C 241
 DB 266 C 266

RESULT 6
 HPCBRND3 321 bp ss-RNA linear VRL 19-APR-1994
 LOCUS HPCBRND3
 DEFINITION Hepatitis C virus isolate D3 5' untranslated region and polyprotein
 mRNA, 5' end.
 ACCESSION M84842
 VERSION M84842.1 GI:329889
 KEYWORDS
 SOURCE Hepatitis C virus (individual isolate D3) cDNA to genomic RNA.
 ORGANISM Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE 1 (bases 1 to 321)
 Buhr, J., Purcell, R.H. and Miller, R.H.
 Sequence analysis of the 5' noncoding region of hepatitis C virus
 Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
 MEDLINE 92279243
 PUBMED 1317578

FEATURES
 source
 1. 321
 /organism="Hepatitis C virus"
 /isolate="D3"
 /db_xref="taxon:11103"
 1. 282
 283..321
 /gene="polyprotein"
 283..321
 /partial
 /gene="polyprotein"
 /codon_start=1
 /product="polyprotein"
 /protein_id="AAA45679.1"
 /db_xref="GI:329889"
 /translation="MSTNPKPQRTKR"

BASE COUNT 69 a 91 c 97 g 64 t
 ORIGIN

Query Match 100.0%; Score 241; DB 14; Length 321;
 Best Local Similarity 100.0%; Pred. No. 6.5e-60;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAACGCTCTACCCATGCGGCTTATGATGATGCTGTCGACGCTCCAGAGACCCCT 60
 DB 9 GCAGAAACGCTCTACCCATGCGGCTTATGATGATGCTGTCGACGCTCCAGAGACCCCT 68
 OY 61 CCCGGAGAGCCATGCTGCTGCGGAGACCGGTGATGACCGGATGCCAGAGACGACC 120
 DB 69 CCCGGAGAGCCATGCTGCTGCGGAGACCGGTGATGACCGGATGCCAGAGACGACC 128

241	C	241
249	-	
249	C	249

ACCESSION M84849
 VERSION M84849.1 GI:329901
 KEYWORDS
 polyprotein.
 Hepatitis C virus (individual isolate HK5) cDNA to genomic RNA.

ORGANISM Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.

REFERENCE 1 (bases 1 to 321)
 AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
 TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
 MEDLINE 92279243
 PUBMED 1317578

FEATURES
 source Location/Qualifiers
 1..321
 /organism="Hepatitis C virus"
 /isolate="HK5"
 /db_xref="taxon:11103"
 5'UTR
 gene 1..282
 283..321
 /gene="polyprotein"
 /partial
 /gene="polyprotein"
 /codon_start=1
 /product="polyprotein"
 /protein_id="AA45685.1"
 /db_xref="GI:329902"
 /translation="MSTNPKPQRKTR"
 BASE COUNT 69 a 91 c 97 g 64 t
 ORIGIN

Query Match 100.0%; Score 241; DB 14; Length 321;
 Best Local Similarity 100.0%; Pred. No. 6,5e-60;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGCTTAGTATGATGTCGTGAGCCTCCAGAGACCCCCCT 60
 Db 9 GCAGAAAGCGTCTAGCCATGCGCTTAGTATGATGTCGTGAGCCTCCAGAGACCCCCCT 68
 QY 61 CCCGGAGAGCCATAGTGTCTGCGGAACCGGTAGTACACCGGAATTGCCAGAGACAC 120
 Db 69 CCCGGAGAGCCATAGTGTCTGCGGAACCGGTAGTACACCGGAATTGCCAGAGACAC 128
 QY 121 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGAGATTGGCGTGCCTCCCGAGAC 180
 Db 129 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGAGATTGGCGTGCCTCCCGAGAC 188
 QY 181 TGTAGCCGAGTACTGTGGTGGCGAAAGCGCTTGTGCTCTGCTGATAGGGTCTTG 240
 Db 189 TGTAGCCGAGTACTGTGGTGGCGAAAGCGCTTGTGCTCTGCTGATAGGGTCTTG 248
 QY 241 C 241
 Db 249 C 249

RESULT 10
 HPCRNP10 321 bp ss-RNA linear VRL 19-APR-1994
 LOCUS Hepatitis C virus isolate P10 5' untranslated region and
 DEFINITION polyprotein mRNA, 5' end.
 ACCESSION M84855
 VERSION M84855.1 GI:329903
 KEYWORDS polyprotein.
 SOURCE Hepatitis C virus (individual isolate P10) cDNA to genomic RNA.
 ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.

REFERENCE 1 (bases 1 to 321)
 AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
 TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
 MEDLINE 92279243
 PUBMED 1317578

FEATURES
 source Location/Qualifiers
 1..321

ORGANISM="Hepatitis C virus"
 /isolate="P10"
 /db_xref="taxon:11103"
 5'UTR
 gene 1..282
 283..321
 /gene="polyprotein"
 /partial
 /gene="polyprotein"
 /codon_start=1
 /product="polyprotein"
 /protein_id="AA45686.1"
 /db_xref="GI:329904"
 /translation="MSTNPKPQRKTR"
 BASE COUNT 69 a 91 c 97 g 64 t
 ORIGIN

Query Match 100.0%; Score 241; DB 14; Length 321;
 Best Local Similarity 100.0%; Pred. No. 6,5e-60;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGCTTAGTATGATGTCGTGAGCCTCCAGAGACCCCCCT 60
 Db 9 GCAGAAAGCGTCTAGCCATGCGCTTAGTATGATGTCGTGAGCCTCCAGAGACCCCCCT 68
 QY 61 CCCGGAGAGCCATAGTGTCTGCGGAACCGGTAGTACACCGGAATTGCCAGAGACAC 120
 Db 69 CCCGGAGAGCCATAGTGTCTGCGGAACCGGTAGTACACCGGAATTGCCAGAGACAC 128
 QY 121 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGAGATTGGCGTGCCTCCCGAGAC 180
 Db 129 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGAGATTGGCGTGCCTCCCGAGAC 188
 QY 181 TGTAGCCGAGTACTGTGGTGGCGAAAGCGCTTGTGCTCTGCTGATAGGGTCTTG 240
 Db 189 TGTAGCCGAGTACTGTGGTGGCGAAAGCGCTTGTGCTCTGCTGATAGGGTCTTG 248
 QY 241 C 241
 Db 249 C 249

RESULT 11
 HPCRNAS45 321 bp ss-RNA linear VRL 19-APR-1994
 LOCUS Hepatitis C virus isolate S45 5' untranslated region and
 DEFINITION polyprotein mRNA, 5' end.
 ACCESSION M84840
 VERSION M84840.1 GI:329907
 KEYWORDS polyprotein.
 SOURCE Hepatitis C virus (individual isolate S45) cDNA to genomic RNA.
 ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.

REFERENCE 1 (bases 1 to 321)
 AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
 TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
 MEDLINE 92279243
 PUBMED 1317578

FEATURES
 source Location/Qualifiers
 1..321
 /organism="Hepatitis C virus"
 /isolate="S45"
 /db_xref="taxon:11103"
 5'UTR
 gene 1..282
 283..321
 /gene="polyprotein"
 /partial
 /gene="polyprotein"
 /codon_start=1
 /product="polyprotein"

/protein_id="AA45688.1"
/db_xref="GI:329908"
/translation="MSTNPKPQRTNR"
BASE COUNT 67 a 93 c 97 g 64 t
ORIGIN

Query Match 100.0%; Score 241; DB 14; Length 321;
Best Local Similarity 100.0%; Pred. No. 6.5e-60;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GCAGAAAGCGTCTAGCCATGCGGCTGTGATGAGTGTGTCGACAGCTCCAGACCCCCCT 60
|||||
9 GCAGAAAGCGTCTAGCCATGCGGCTGTGATGAGTGTGTCGACAGCTCCAGACCCCCCT 68
|||||

61 CCCGGAGAGCCATAGTGTCTGCGGACCGGTGAGTACACCGGAATTGCCAGAGACCC 120
|||||
69 CCCGGAGAGCCATAGTGTCTGCGGACCGGTGAGTACACCGGAATTGCCAGAGACCC 128
|||||

121 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 180
|||||
129 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 188
|||||

181 TGTAGCCGATAGTGTGTCGCGAAAGCGCTTGTGCTACTGCTCATAGGCTTG 240
|||||
189 TGTAGCCGATAGTGTGTCGCGAAAGCGCTTGTGCTACTGCTCATAGGCTTG 248
|||||

241 C 241
249 C 249

RESULT 12
HPCRNAS9
LOCUS HPCRNAS9 321 bp ss-RNA linear VRL 19-APR-1994
DEFINITION Hepatitis C virus isolate 5' S9 untranslated region and polyprotein
M84838
VERSION M84838.1 GI:329913
KEYWORDS polyprotein.
SOURCE Hepatitis C virus (individual isolate S9) cDNA to genomic RNA.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE 1 (bases 1 to 321)
AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
MEDLINE 92279243
PUBMED 1317578

FEATURES
source Location/Qualifiers
1..321
/organism="Hepatitis C virus"
/isolate="S9"
/db_xref="taxon:11103"
1..282
283..321
/gene="polyprotein"
283..321
/partial
/gene="polyprotein"
/codon_start=1
/product="polyprotein"
/protein_id="AA45691.1"
/db_xref="GI:329914"
/translation="MSTNPKPQRTNR"
BASE COUNT 69 a 91 c 97 g 64 t
ORIGIN

Query Match 100.0%; Score 241; DB 14; Length 321;
Best Local Similarity 100.0%; Pred. No. 6.5e-60;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GCAGAAAGCGTCTAGCCATGCGGCTGTGATGAGTGTGTCGACAGCTCCAGACCCCCCT 60
|||||

|||||
9 GCAGAAAGCGTCTAGCCATGCGGCTGTGATGAGTGTGTCGACAGCTCCAGACCCCCCT 68
|||||

61 CCCGGAGAGCCATAGTGTCTGCGGAAACCGGTGAGTACACCGGAATTGCCAGAGACCC 120
|||||
69 CCCGGAGAGCCATAGTGTCTGCGGAAACCGGTGAGTACACCGGAATTGCCAGAGACCC 128
|||||

121 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 180
|||||
129 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 188
|||||

181 TGTAGCCGATAGTGTGTCGCGAAAGCGCTTGTGCTACTGCTCATAGGCTTG 240
|||||
189 TGTAGCCGATAGTGTGTCGCGAAAGCGCTTGTGCTACTGCTCATAGGCTTG 248
|||||

241 C 241
249 C 249

RESULT 13
HPCRNASW2
LOCUS HPCRNASW2 321 bp ss-RNA linear VRL 19-APR-1994
DEFINITION Hepatitis C virus isolate SW2 5' untranslated region and polyprotein mRNA, 5' end.
M84863
VERSION M84863.1 GI:329920
KEYWORDS polyprotein.
SOURCE Hepatitis C virus (individual isolate SW2) cDNA to genomic RNA.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE 1 (bases 1 to 321)
AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
MEDLINE 92279243
PUBMED 1317578

FEATURES
source Location/Qualifiers
1..321
/organism="Hepatitis C virus"
/isolate="SW2"
/db_xref="taxon:11103"
1..282
283..321
/gene="polyprotein"
283..321
/partial
/gene="polyprotein"
/codon_start=1
/product="polyprotein"
/protein_id="AA45694.1"
/db_xref="GI:329921"
/translation="MSTNPKPQRTNR"
BASE COUNT 69 a 91 c 97 g 64 t
ORIGIN

Query Match 100.0%; Score 241; DB 14; Length 321;
Best Local Similarity 100.0%; Pred. No. 6.5e-60;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GCAGAAAGCGTCTAGCCATGCGGCTGTGATGAGTGTGTCGACAGCTCCAGACCCCCCT 60
|||||

9 GCAGAAAGCGTCTAGCCATGCGGCTGTGATGAGTGTGTCGACAGCTCCAGAGACCC 68
|||||

61 CCCGGAGAGCCATAGTGTCTGCGGAAACCGGTGAGTACACCGGAATTGCCAGAGACCC 120
|||||
69 CCCGGAGAGCCATAGTGTCTGCGGAAACCGGTGAGTACACCGGAATTGCCAGAGACCC 128
|||||

121 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 180
|||||
129 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTCCCGCGAGAC 188
|||||

RESULT 14
LOCUS HPCRNAUS3 321 bp ss-RNA linear VRL 19-APR-1994
DEFINITION Hepatitis C virus isolate US3 5' untranslated region and
polypeptide mRNA, 5' end.
ACCESSION M84857.1 GI:329933
VERSION M84857
KEYWORDS polypeptide.
SOURCE Hepatitis C virus (individual isolate US3) cDNA to genomic RNA.
ORGANISM
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE 1 (bases 1 to 321)
AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
MEDLINE 92279243
PUBMED 1317578
FEATURES
source
Location/Qualifiers
1..321
/organism="Hepatitis C virus"
/isolate="US3"
/db_xref="taxon:11103"
1..282
/gene="polyprotein"
283..321
/partial
/gene="polyprotein"
/codon_start=1
/product="polyprotein"
/protein_id="AA45699.1"
/db_xref="GI:329934"
/translation="MSTNPKPQRRTKR"
BASE COUNT 69 a 91 c 97 g 64 t
ORIGIN
Query Match 100.0%; Score 241; DB 14; Length 321;
Best Local Similarity 100.0%; Pred. No. 6.5e-60;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAGAAAGCGCTAGTGGTGGCGAAGGCGCTTGTGATGCTGATAGGCTGCTG 240
DB 9 GCAGAAAGCGCTAGTGGTGGCGAAGGCGCTTGTGATGCTGATAGGCTGCTG 68
QY 61 CCCGGAGAGCCATAGTGGTGGCGAAGGCGCTTGTGATGCTGATAGGCTGCTG 120
DB 69 CCCGGAGAGCCATAGTGGTGGCGAAGGCGCTTGTGATGCTGATAGGCTGCTG 128
QY 121 GGGTCCTTCTTGGATCAACCCGCTCAATGCTGAGATTGGCGGCGGAGAC 180
DB 129 GGGTCCTTCTTGGATCAACCCGCTCAATGCTGAGATTGGCGGCGGAGAC 188
QY 181 TGCTAGCCGAGTGTGGTGGCGAAGGCGCTTGTGATGCTGATAGGCTGCTG 240
DB 189 TGCTAGCCGAGTGTGGTGGCGAAGGCGCTTGTGATGCTGATAGGCTGCTG 248
QY 241 C 241
DB 249 C 249

RESULT 15
HPCRNAUS6

LOCUS HPCRNAUS6 321 bp ss-RNA linear VRL 19-APR-1994
DEFINITION Hepatitis C virus isolate US6 5' untranslated region and
polypeptide, 5' end.
ACCESSION M84830.1 GI:329935
VERSION M84830
KEYWORDS polypeptide.
SOURCE Hepatitis C virus (individual isolate US6) cDNA to genomic RNA.
ORGANISM
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE 1 (bases 1 to 321)
AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
MEDLINE 92279243
PUBMED 1317578
FEATURES
source
Location/Qualifiers
1..321
/organism="Hepatitis C virus"
/isolate="US6"
/db_xref="taxon:11103"
1..282
/gene="polyprotein"
283..321
/partial
/gene="polyprotein"
/codon_start=1
/product="polyprotein"
/protein_id="AA45700.1"
/db_xref="GI:329936"
/translation="MSTNPKPQRRTKR"
BASE COUNT 69 a 91 c 97 g 64 t
ORIGIN
Query Match 100.0%; Score 241; DB 14; Length 321;
Best Local Similarity 100.0%; Pred. No. 6.5e-60;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAGAAAGCGCTAGTGGTGGCGAAGGCGCTTGTGATGCTGATAGGCTGCTG 60
DB 9 GCAGAAAGCGCTAGTGGTGGCGAAGGCGCTTGTGATGCTGATAGGCTGCTG 68
QY 61 CCCGGAGAGCCATAGTGGTGGCGAAGGCGCTTGTGATGCTGATAGGCTGCTG 120
DB 69 CCCGGAGAGCCATAGTGGTGGCGAAGGCGCTTGTGATGCTGATAGGCTGCTG 128
QY 121 GGGTCCTTCTTGGATCAACCCGCTCAATGCTGAGATTGGCGGCGGAGAC 180
DB 129 GGGTCCTTCTTGGATCAACCCGCTCAATGCTGAGATTGGCGGCGGAGAC 188
QY 181 TGCTAGCCGAGTGTGGTGGCGAAGGCGCTTGTGATGCTGATAGGCTGCTG 240
DB 189 TGCTAGCCGAGTGTGGTGGCGAAGGCGCTTGTGATGCTGATAGGCTGCTG 248
QY 241 C 241
DB 249 C 249

Search completed: July 10, 2003, 19:20:43
Job time : 2092.31 secs

